7A catalogue of molecular targets for kidney function from genetic analyses of a million individuals

Matthias Wuttke*[1], Yong Li*[1], Man Li*[2], Karsten B. Sieber*[3], Mary F. Feitosa*[4], Mathias Gorski*[5,6], Adrienne Tin[7,8], Lihua Wang[4], Audrey Y. Chu[9], Anselm Hoppmann[1], Holger Kirsten[10,11], Ayush Giri[12,13,14], Jin-Fang Chai[15], Gardar Sveinbjornsson[16], Bamidele O. Tayo[17], Teresa Nutile[18], Christian Fuchsberger[19], Jonathan Marten[20], Massimiliano Cocca[21], Sahar Ghasemi[22,23], Yizhe Xu[2], Katrin Horn[10,11], Damia Noce[19], Peter J. van der Most[24], Sanaz Sedaghat[25], Zhi Yu[7,26], Masato Akiyama[27,28], Saima Afaq[29,30], Tarunveer S. Ahluwalia[31], Peter Almgren[32], Najaf Amin[25], Johan Ärnlöv[33,34], Stephan J.L. Bakker[35], Nisha Bansal[36,37], Daniela Baptista[38], Sven Bergmann[39,40,41], Mary L. Biggs[42,43], Ginevra Biino[44], Michael Boehnke[45,46], Eric Boerwinkle[47], Mathilde Boissel[48], Erwin P. Bottinger[49,50], Thibaud S. Boutin[20], Hermann Brenner[51,52], Marco Brumat[53], Ralph Burkhardt[11,54,55], Adam S. Butterworth[56,57], Eric Campana[53], Archie Campbell[58], Harry Campbell[59], Mickaël Canouil[48], Robert J. Carroll[60], Eulalia Catamo[21], John C. Chambers[29,61,62,63,269], Miao-Ling Chee[64], Miao-Li Chee[64], Xu Chen[65], Ching-Yu Cheng[64,66,67], Yurong Cheng[1], Kaare Christensen[68], Renata Cifkova[69,70], Marina Ciullo[18,71], Maria Pina Concas[21], James P. Cook[72], Josef Coresh[7], Tanguy Corre[39,40,73], Daniele Cusi[74,75], Sala Cinzia Felicita[76], John Danesh[77], E. Warwick Daw[4], Martin H. de Borst[35], Alessandro De Grandi[19], Renée de Mutsert[78], Aiko P.J. de Vries[79], Frauke Degenhardt[80], Graciela Delgado[81], Ayse Demirkan[25], Emanuele di Angelantonio[82,83,84], Katalin Dittrich[85,86], Jasmin Divers[87], Rajkumar Dorajoo[88], Kai-Uwe Eckardt[89,90], Georg Ehret[38], Paul Elliott[91,92,93,94], Karlhans Endlich[23,95], Michele K. Evans[96], Janine F. Felix[25,97,98], Valencia Hui Xian Foo[64], Oscar H. Franco[25,99], Andre Franke[80], Barry I. Freedman[100], Sandra Freitag-Wolf[101], Yechiel Friedlander[102], Philippe Froguel[48,103], Ron T. Gansevoort[35], He Gao[91], Paolo Gasparini[21,53], J. Michael Gaziano[104], Vilmantas Giedraitis[105], Christian Gieger[106,107,108], Giorgia Girotto[21,53], Franco Giulianini[109], Martin Gögele[19], Scott D. Gordon[110], Daniel F. Gudbjartsson[16], Vilmundur Gudnason[111,112], Toomas Haller[113], Pavel Hamet[114,115], Tamara B. Harris[116], Catharina A. Hartman[117], Caroline Hayward[20], Jacklyn N. Hellwege[14,118,119], Chew-Kiat Heng[120,121,122], Andrew A. Hicks[19], Edith Hofer[123,124], Wei Huang[125,126], Nina Hutri-Kähönen[127,128], Shih-Jen Hwang[129,130], M. Arfan Ikram[25], Olafur S. Indridason[131], Erik Ingelsson[132,133,134], Marcus Ising[135], Vincent W.V. Jaddoe[25,97,98], Johanna Jakobsdottir[111], Jost B. Jonas[136,137], Peter K. Joshi[59], Navya Shilpa Josyula[138], Bettina Jung[5], Mika Kähönen[139,140], Yoichiro Kamatani[27,141], Candace M. Kammerer[142], Masahiro Kanai[27,143], Mika Kastarinen[144], Shona M. Kerr[20], Chiea-Chuen Khor[64,88], Wieland Kiess[11,85,86], Marcus E. Kleber[81], Wolfgang Koenig[145], Jaspal S. Kooner[62,63,146,147], Antje Körner[11,85,86], Peter Kovacs[148], Aldi T. Kraja[4], Alena Krajcoviechova[69,70], Holly Kramer[17,149], Bernhard K. Krämer[81], Florian Kronenberg[150], Michiaki Kubo[151], Brigitte Kühnel[106], Mikko Kuokkanen[152,153], Johanna Kuusisto[154,155], Markku Laakso[154,155], Martina La Bianca[21], Leslie A. Lange[156], Carl D. Langefeld[87], Jeannette Jen-Mai Lee[15], Benjamin Lehne[29], Terho Lehtimäki[140], Wolfgang Lieb[157], Lifelines Cohort Study[158], Su-Chi Lim[15,159], Lars Lind[160], Cecilia M. Lindgren[161,162], Jun Liu[25], Jianjun Liu[88,163], Markus Loeffler[10,11], Ruth J.F. Loos[49,164], Susanne Lucae[135], Mary Ann Lukas[165], Leo-Pekka Lyytikäinen[140], Reedik Mägi[113], Patrik K.E. Magnusson[65], Anubha Mahajan[166,167], Nicholas G. Martin[110], Jade Martins[168], Winfried März[169,170,171], Deborah Mascalzoni[19], Koichi Matsuda[172], Christa

Meisinger[107,173,174,175], Thomas Meitinger[176,177], Olle Melander[178], Andres Metspalu[113], Evgenia K. Mikaelsdottir[16], Yuri Milaneschi[179], Kozeta Miliku[25,97,98], Million Veteran Program[180], Pashupati P. Mishra[140], Karen L. Mohlke[181], Nina Mononen[140], Grant W. Montgomery[182], Dennis O. Mook-Kanamori[78,183], Josyf C. Mychaleckyj[184], Girish N. Nadkarni[49,185], Mike A. Nalls[186,187], Matthias Nauck[23,188], Kjell Nikus[189,190], Boting Ning[191], Ilja M. Nolte[24], Raymond Noordam[192], Jeffrey O'Connell[193], Michelle L. O'Donoghue[194,195], Isleifur Olafsson[196], Albertine J. Oldehinkel[117], Marju Orho-Melander[32], Willem H. Ouwehand[77], Sandosh Padmanabhan[197], Nicholette D. Palmer[198], Runolfur Palsson[112,131], Brenda W.J.H. Penninx[179], Thomas Perls[199], Markus Perola[200,201], Mario Pirastu[202], Nicola Pirastu[59], Giorgio Pistis[203], Anna I. Podgornaia[9], Ozren Polasek[204,205], Belen Ponte[206], David J. Porteous[58,207], Tanja Poulain[11], Peter P. Pramstaller[19], Michael H. Preuss[49], Bram P. Prins[56], Michael A. Province[4], Ton J. Rabelink[79,208], Laura M. Raffield[181], Olli T. Raitakari[209,210], Dermot F. Reilly[9], Rainer Rettig[211], Myriam Rheinberger[5], Kenneth M. Rice[43], Paul M. Ridker[109,212], Fernando Rivadeneira[25,213], Federica Rizzi[214,215], David J. Roberts[82,216], Antonietta Robino[21], Peter Rossing[31], Igor Rudan[59], Rico Rueedi[39,40], Daniela Ruggiero[18,71], Kathleen A. Ryan[217], Yasaman Saba[218], Charumathi Sabanayagam[64,66], Veikko Salomaa[200], Erika Salvi[214,219], Kai-Uwe Saum[51], Helena Schmidt[220], Reinhold Schmidt[123], Ben Schöttker[51,52], Christina-Alexandra Schulz[32], Nicole Schupf[221,222,223], Christian M. Shaffer[60], Yuan Shi[64,66], Albert V. Smith[224], Blair H. Smith[225], Nicole Soranzo[226], Cassandra N. Spracklen[181], Konstantin Strauch[227,228], Heather M. Stringham[45,46], Michael Stumvoll[229], Per O. Svensson[230,231], Silke Szymczak[101], E-Shyong Tai[15,232,233], Salman M. Tajuddin[96], Nicholas Y. Q. Tan[64], Kent D. Taylor[234], Andrej Teren[11,235], Yih-Chung Tham[64], Joachim Thiery[11,54], Chris H.L. Thio[236], Hauke Thomsen[237], Gudmar Thorleifsson[16], Daniela Toniolo[76], Anke Tönjes[229], Johanne Tremblay[114,238], Ioanna Tzoulaki[91,239], André G. Uitterlinden[213], Simona Vaccargiu[202], Rob M. van Dam[15,163,240], Pim van der Harst[241,242,243], Cornelia M. van Duijn[25], Digna R. Velez Edward[12,13,14], Niek Verweij[241], Suzanne Vogelezang[25,97,98], Uwe Völker[23,244], Peter Vollenweider[245], Gerard Waeber[245], Melanie Waldenberger[106,107,145], Lars Wallentin[246,247], Ya Xing Wang[137], Chaolong Wang[88,248], Dawn M. Waterworth[3], Wen Bin Wei[249], Harvey White[250], John B. Whitfield[110], Sarah H. Wild[251], James F. Wilson[20,59], Mary K. Wojczynski[4], Charlene Wong[67], Tien-Yin Wong[64,66,67], Liang Xu[137], Qiong Yang[191], Masayuki Yasuda[64,252], Laura M. Yerges-Armstrong[3], Weihua Zhang[62,91], Alan B. Zonderman[96], Jerome I. Rotter[234,253,254], Murielle Bochud[73], Bruce M. Psaty[42,255], Veronique Vitart[20], James G. Wilson[256], Abbas Dehghan[29,91], Afshin Parsa[257,258], Daniel I. Chasman[109,212], Kevin Ho[259,260], Andrew P. Morris[72,166], Olivier Devuyst[261], Shreeram Akilesh[37,262], Sarah A. Pendergrass[263], Xueling Sim[15], Carsten A. Böger[5,264], Yukinori Okada[265,266], Todd L. Edwards[14,267], Harold Snieder[236], Kari Stefansson[16], Adriana M. Hung[14,268], Iris M. Heid**[6], Markus Scholz**[10,11], Alexander Teumer**[22,23], Anna Köttgen**[1,7], Cristian Pattaro**[19]

* Indicates joint contribution

** Indicates joint oversight

† Indicates corresponding author

Authors for Correspondence:

Anna Köttgen, MD MPH
Institute of Genetic Epidemiology
Medical Center - University of Freiburg
Hugstetter Str. 49, 79106 Freiburg, Germany
+49 (0)761 270-78050
anna.koettgen@uniklinik-freiburg.de

Cristian Pattaro PhD
Eurac Research
Institute for Biomedicine
Via Galvani 31, 39100 Bolzano, Italy
+39 0471 055 527
cristian.pattaro@eurac.edu

Author affiliations

1 Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Centre - University of Freiburg, Freiburg, Germany
2 Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City, USA
3 Target Sciences - Genetics, GlaxoSmithKline, Collegeville (Pennsylvania), USA
4 Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis (Missouri), USA
5 Department of Nephrology, University Hospital Regensburg, Regensburg, Germany
6 Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany
7 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore (Maryland), USA
8 Epidemiology and Clinical Research, Welch Centre for Prevention, Baltimore (Maryland), USA
9 Genetics, Merck Sharp & Dohme Corp, Kenilworth, New Jersey, USA
10 Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany
11 LIFE Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany
12 Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt University Medical Centre, Nashville (TN), USA
13 Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, USA
14 Department of Veteranâ€™s Affairs, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA
15 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore
16 deCODE Genetics, Amgen Inc., Reykjavik, Iceland
17 Department of Public Health Sciences, Loyola University Chicago, Maywood (Illinois), USA
18 Institute of Genetics and Biophysics "Adriano Buzzati-Traverso" - CNR, Naples, Italy
19 Eurac Research, Institute for Biomedicine (affiliated to the University of Lübeck), Bolzano, Italy
20 Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
21 Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy
22 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
23 DZHK (German Center for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany
24 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
25 Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands
26 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore (Maryland), USA
27 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
28 Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
29 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial College London, London, UK
30 Institute of Public health & social sciences, Khyber Medical University, Pakistan
31 Steno Diabetes Centre Copenhagen, Gentofte, Denmark
32 Diabetes and Cardiovascular Disease - Genetic Epidemiology, Department of Clincial Sciences in Malmö, Lund University, Malmö, Sweden
33 Department of Neurobiology, Care Sciences and Society, Division of Family Medicine and Primary Care, Karolinska Institutet, Stockholm, Sweden
34 School of Health and Social Studies, Dalarna University, Sweden
35 Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
36 Division of Nephrology, University of Washington, Seattle (Washington), USA
37 Kidney Research Institute, University of Washington, Seattle (Washington), USA
38 Cardiology, Geneva University Hospitals, Geneva, Switzerland
39 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
40 Swiss Institute of Bioinformatics, Lausanne, Switzerland
41 Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa
42 Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Service, University of Washington, Seattle (Washington), USA
43 Department of Biostatistics, University of Washington, Seattle (Washington), USA
44 Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy
45 Department of Biostatistics, University of Michigan, Ann Arbor, USA
46 Centre for Statistical Genetics, University of Michigan, USA
47 Human Genetics Centre, University of Texas Health Science Centre, Houston (Texas), USA
48 CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille, France
49 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York (New York), USA
50 Digital Health Centre, Hasso Plattner Institute and University of Potsdam, Potsdam, Germany
51 Division of Clinical Epidemiology and Aging Research, German Cancer Research Centre (DKFZ), Heidelberg, Germany
52 Network Aging Research, University of Heidelberg, Heidelberg, Germany
53 University of Trieste, Department of Medicine, Surgery and Health Sciences, Trieste, Italy
54 Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany
55 Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany
56 MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
57 National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK

58 Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
59 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
60 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville (Tennessee), USA
61 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore
62 Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK
63 Imperial College Healthcare NHS Trust, Imperial College London, London, UK
64 Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore
65 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
66 Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore
67 Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore
68 Unit of Epidemiology, Biostatistics and Biodemography, Department of Public Health, Southern Denmark University, Odense, Denmark
69 Centre for Cardiovascular Prevention, First Faculty of Medicine, Department of Medicine, Charles University in Prague, Prague, Czech Republic
70 Thomayer Hospital, Prague, Czech Republic
71 IRCCS Neuromed, Pozzilli, Italy
72 Department of Biostatistics, University of Liverpool, Liverpool, UK
73 Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland
74 Institute of Biomedical Technologies, Italy National Research Council, Bresso (Milano), Italy
75 Bio4Dreams - business nursery for life sciences, Bresso (Milano), Italy
76 San Raffaele Research Institute, Milano, Italy
77 University of Cambridge, Cambridge, UK
78 Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands
79 Section of Nephrology, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands
80 Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
81 5th Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
82 Blood and Transplant Research Unit in Donor Health and Genomics, National Institute of Health Research, Cambridge, UK
83 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
84 NHS Blood and Transplant, Cambridge, UK
85 Department of Women and Child Health, Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany
86 Centre for Pediatric Research, University of Leipzig, Leipzig, Germany
87 Public Health Sciences - Biostatistics, Wake Forest School of Medicine, Winston-Salem (North Carolina), USA
88 Genome Institute of Singapore, Agency for Science Technology and Research, Singapore, Singapore
89 Medical Department, Division of Nephrology and Internal Intensive Care Medicine CVK/CCM, Charité Universität Medizin Berlin, Germany
90 Department of Nephrology and Hypertension, Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and University Hospital Erlangen, Germany
91 Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK
92 Imperial College NIHR Biomedical Research Centre, Imperial College London, London, UK
93 Dementia Research Institute, Imperial College London, London, UK
94 Health Data Research UK-London, London, UK
95 Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany
96 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Baltimore (Maryland), USA
97 The Generation R Study Group, Erasmus University Medical Center, Rotterdam, The Netherlands
98 Department of Pediatrics, Erasmus University Medical Center, Rotterdam, The Netherlands
99 Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
100 Internal Medicine - Section on Nephrology, Wake Forest School of Medicine, Winston-Salem (North Carolina), USA
101 Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Kiel
102 School of Public Health and Community Medicine, Hebrew University of Jerusalem, Jerusalem, Israel
103 Department of Genomics of Common Disease, Imperial College London, London, UK
104 Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies Program, VA Boston Healthcare System, Boston (Massachusetts), USA
105 Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala, Sweden
106 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
107 Institute of Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
108 German Center for Diabetes Research (DZD), Neuherberg, Germany
109 Division of Preventive Medicine, Brigham and Women's Hospital, Boston, USA
110 QIMR Berghofer Medical Research Institute, Brisbane, Australia
111 Icelandic Heart Association, Kopavogur, Iceland
112 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland
113 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
114 Montreal University Hospital Research Centre, CHUM, Montreal, Canada

115 Medpharmgene, Montreal, Canada
116 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda (Maryland), USA
117 Interdisciplinary Centre Psychopathology and Emotion regulation (ICPE), University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
118 Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, USA
119 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, USA
120 Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
121 National University Children's Medical Institute, Khoo Teck Puat, Singapore
122 National University Health System, Singapore, Singapore
123 Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria
124 Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria
125 Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Centre, Shanghai, China
126 Shanghai Industrial Technology Institute, Shanghai, China
127 Department of Pediatrics, Tampere University Hospital, Tampere, Finland
128 Department of Pediatrics, Faculty of Medicine and Life Sciences, University of Tampere, Finland
129 NHLBI's Framingham Heart Study, Framingham (Massachusetts), USA
130 The Centre for Population Studies, NHLBI, Framingham (Massachusetts), USA
131 Division of Nephrology, Internal Medicine Services, Landspitaliâ€"The National University Hospital of Iceland, Reykjavik, Iceland
132 Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, USA
133 Stanford Cardiovascular Institute, Stanford University, USA
134 Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
135 Max Planck Institute of Psychiatry, Munich, Germany
136 Department of Ophthalmology, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany
137 Beijing Institute of Ophthalmology, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing Tongren Hospital, Capital Medical University, Beijing, China
138 Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, USA
139 Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
140 Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
141 Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
142 Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh (Pennsylvania), USA
143 Department of Biomedical Informatics, Harvard Medical School, Boston, USA
144 Kuopio University Hospital, Kuopio, Finland
145 DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany
146 MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK
147 National Heart and Lung Institute, Imperial College London, London, UK
148 Integrated Research and Treatment Centre Adiposity Diseases, University of Leipzig, Leipzig, Germany
149 Division of Nephrology and Hypertension, Loyola University Chicago, USA
150 Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria
151 RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
152 The Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland
153 Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland
154 University of Eastern Finland, Kuopio, Finland
155 Kuopio University Hospital, Finland
156 Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver - Anschutz Medical Campus, Aurora (Colorado), USA
157 Institute of Epidemiology and Biobank Popgen, Kiel University, Kiel, Germany
158 Lifelines Cohort Study
159 Diabetes Centre, Khoo Teck Puat Hospital, Singapore, Singapore
160 Cardiovascular Epidemiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
161 Nuffield Department of Medicine, University of Oxford, Oxford, UK
162 Broad Institute of Harvard and MIT, USA
163 Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore
164 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York (New York), USA
165 Target Sciences - Genetics, GlaxoSmithKline, Albuquerque (New Mexico), USA
166 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
167 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, UK
168 Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany
169 Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany
170 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Austria
171 Medical Clinic V, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
172 Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

173 Independent Clinical Epidemiology Research Group, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany
174 Ludwig-Maximilians-Universität München, Munich, Germany
175 Epidemiology, UNIKA-T Augsburg, Augsburg, Germany
176 Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
177 Institute of Human Genetics, Technische Universität München, Munich, Germany
178 Hypertension and Cardiovascular Disease, Department of Clincial Sciences Malmö, Lund University, Malmö, Sweden
179 Department of Psychiatry, VU University Medical Centre, Amsterdam, The Netherlands
180 Department of Veterans Affairs. Office of Research and Development, Washington, DC, USA
181 Department of Genetics, University of North Carolina, Chapel Hill (North Carolina), USA
182 University of Queensland, St Lucia, Australia
183 Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, The Netherlands
184 Centre for Public Health Genomics, University of Virginia, Charlottesville (Virginia), USA
185 Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York (New York), USA
186 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda (Maryland), USA
187 Data Tecnica International, Glen Echo (Maryland), USA
188 Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
189 Department of Cardiology, Heart Centre, Tampere University Hospital, Tampere, Finland
190 Department of Cardiology, Finnish Cardiovascular Research Centre - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
191 Department of Biostatistics, Boston University School of Public Health, Boston (Massachusetts), USA
192 Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands
193 University of Maryland School of Medicine, Baltimore, USA
194 Cardiovascular Division, Brigham and Women's Hospital, Boston, USA
195 TIMI Study Group, USA
196 Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland
197 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
198 Biochemistry, Wake Forest School of Medicine, Winston-Salem (North Carolina), USA
199 Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston (Massachusetts), USA
200 National Institute for Health and Welfare, Helsinki, Finland
201 The Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland
202 Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li Punti (Sassari), Italy
203 Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland
204 Faculty of Medicine, University of Split, Split, Croatia
205 Gen-info Ltd, Zagreb, Croatia
206 Service de Néphrologie, Geneva University Hospitals, Geneva, Switzerland
207 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
208 Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Centre, Leiden, The Netherlands
209 Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
210 Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
211 Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany
212 Harvard Medical School, USA
213 Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands
214 Department of Health Sciences, University of Milan, Milano, Italy
215 ePhood Scientific Unit, ePhood SRL, Milano, Italy
216 NHS Blood and Transplant; BRC Oxford Haematology Theme; Nuffield Division of Clinical Laboratory Sciences; University of Oxford, Oxford, UK
217 Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, USA
218 Molecular Biology and Biochemistry, Gottfried Schatz Research Centre for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria
219 Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milan, Italy
220 Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria
221 Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, USA
222 Gertrude H. Sergievsky Centre, Columbia University Medical Centre, USA
223 Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Centre, New York, USA
224 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavík, Iceland
225 Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK
226 Wellcome Sanger Institute, South Cambridgeshire, UK
227 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
228 Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany
229 Department of Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany
230 Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset (KI SÖS), Stockholm, Sweden
231 Department of Cardiology, Karolinska Institutet, Södersjukhuset (KI SÖS), Stockholm, Sweden
232 Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore
233 Duke-NUS Medical School, Singapore, Singapore

234 The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA
235 Heart Centre Leipzig, Leipzig, Germany
236 Department of Epidemiology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
237 Division of Molecular Genetic Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg, Germany
238 CRCHUM, Montreal, Canada
239 Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece
240 Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, USA
241 Department of Cardiology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
242 Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
243 Durrer Centre for Cardiovascular Research, The Netherlands Heart Institute, Utrecht, The Netherlands
244 Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
245 Internal Medicine, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland
246 Cardiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
247 Uppsala Clinical Research Centre, Uppsala University. Uppsala, Sweden
248 School of Public Health, Tongji Medical School, Huazhong University of Science and Technology, China
249 Beijing Tongren Eye Centre, Beijing Tongren Hospital, Capital Medical University, Beijing, China
250 Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand
251 Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
252 Department of Ophthalmology, Tohoku University Graduate School of Medicine, Japan
253 Department of Pediatrics, Harbor-UCLA Medical Centre, USA
254 Department of Medicine, Harbor-UCLA Medical Centre, Torrance, USA
255 Kaiser Permanente Washington Health Research Institute, Seattle (Washington), USA
256 Department of Physiology and Biophysics, University of Mississippi Medical Centre, Jackson (Mississippi), USA
257 Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, USA
258 Department of Medicine, University of Maryland School of Medicine, Baltimore, USA
259 Kidney Health Research Institute (KHRI), Geisinger, Danville (Pennsylvania), USA
260 Department of Nephrology, Geisinger, Danville (Pennsylvania), USA
261 Institute of Physiology, University of Zurich, Zurich, Switzerland
262 Anatomic Pathology, University of Washington Medical Center, Seattle, USA
263 Geisinger Research, Biomedical and Translational Informatics Institute, Danville (Pennsylvania), USA
264 Department of Nephrology and Rheumatology, Kliniken Südostbayern AG, Regensburg, Germany
265 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Osaka, Japan
266 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan
267 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville (TN), USA
268 Vanderbilt University Medical Centre, Division of Nephrology & Hypertension, Nashville (TN), USA
269 MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, UK

Abstract

Chronic kidney disease is a worldwide public health concern with multi-systemic complications. We performed a trans-ethnic meta-analysis of genome-wide association studies of estimated glomerular filtration rate (eGFR, n=765,348), identifying 308 loci that explained 20% of eGFR heritability. Results were externally replicated (n=280,722) and characterized with the alternative kidney function marker blood urea nitrogen with respect to their kidney function relevance. Pathway and enrichment analyses, including genetically manipulated mice with renal phenotypes, support the kidney as the main target organ. A genetic risk score for low eGFR was significantly associated with clinical diagnosed CKD and related traits among 452,264 independent individuals. Gene expression co-localization analyses across 46 human tissues, including tubulo-interstitial and glomerular kidney compartments, identified 18 kidney-specific prioritized target genes such as *UMOD*, *KNG1*, *MLLT3*, and *GALNTL5*. Fine-mapping highlighted missense driver variants in 10 genes, including several renal transporters, and a kidney-specific regulatory variant in *PDILT* upstream of *UMOD* sharing associations with eGFR, *UMOD* expression, and urinary uromodulin levels. These results provide a comprehensive priority list of molecular targets for translational research.

Chronic kidney disease (CKD) is a major public health issue, with increasing incidence and prevalence worldwide.[1] Its associated burden of disease encompasses metabolic disturbances, end-stage renal disease, and multisystemic complications such as cardiovascular disease.[1-4] CKD is a leading cause of death[5] and shows one of the highest increases in disease-attributable mortality over the last decade.[2] Nevertheless, public and clinical awareness remains low.[3] Moreover, clinical trials in nephrology are still underrepresented compared to other disciplines,[6] resulting in scarce therapeutic options to alter disease progression and high costs for health systems.[7] A major barrier to developing new therapeutics is the limited understanding of the mechanisms underlying kidney function in health and disease, and consequently in a lack of therapeutic targets.

Genome-wide association studies (GWAS) and exome-chip studies of the glomerular filtration rate estimated from serum creatinine (eGFR), the main biomarker to quantify kidney function and define CKD, have shed light on the underlying mechanisms of CKD.[8] Nearly one hundred genetic loci were identified in samples of European[9-15], Asian[16-18], and multiple[19] ancestries. However, similar to other complex traits and diseases, identifying causal genes and molecular mechanisms implicated by genetic associations represents a substantial challenge and has only been successful for few kidney function-associated loci.[20,21] Advanced statistical fine-mapping approaches and newly emerging gene expression data across a wide range of tissues open up new opportunities for prioritizing putative causal variants, effector genes, and target tissues, based on the results from large-scale GWAS meta-analyses.

In addition, the detection of co-expressed genes as well as gene sets, cell-type specific regulatory marks and pathways that are enriched for trait-associated signals is now possible, but requires particularly large GWAS sample sizes. The largest published GWAS meta-analyses of eGFR included up to 140,000 individuals.[9,10,22] However, the identified index variants only explained <4% of the eGFR variance.[9,10] A substantial expansion of study sample size, inclusion of more diverse populations, and more comprehensive coverage of genetic variants promise to identify novel loci, increase the explained variance of eGFR, and detect disease-relevant pathways and co-regulation with other complex traits.

We therefore carried out a trans-ethnic meta-analysis of GWAS of eGFR from 765,348 individuals in order to maximize statistical power to identify novel eGFR-associated loci, which were globally representative loci. Generalizability of results was evaluated through replication in an independent study of 280,722 individuals, for a combined sample size of >1 million participants, and with genetic risk score analyses of clinical diagnoses of kidney disease in an independent sample of 452,264 individuals. Associated loci were characterized through a complementary kidney function marker, blood urea nitrogen (BUN), which was used to prioritize loci with respect to their kidney function relevance. To identify most likely causal variants, genes, and mechanisms, we performed enrichment and network analyses, statistical fine-mapping, and integration of gene expression in kidney and 44 other tissues. The resulting list of functionally relevant variants, genes, tissues and pathways provides a rich resource of potential therapeutic targets to improve CKD treatment and prevention.

Results

Overview

Within the CKD Genetics (CKDGen) Consortium, we established a collaborative, standardized and automated analysis workflow to integrate results from 121 eGFR GWAS of five ancestry groups (**Supplementary Table 1**). Our effort served two objectives (**Supplementary Figure 1**): first, we aimed at identifying novel, globally representative loci for kidney function through meta-analysis of trans-ethnic samples; second, we aimed to understand each locus in depth through complementary computational approaches, including various enrichment analyses, statistical fine-mapping and co-localization with gene expression and protein levels in urine, among European ancestry (EA) individuals, for whom large reference panels on linkage disequilibrium (LD) structure are available.

Identification of 308 loci associated with eGFR through trans-ethnic meta-analysis

In total, the 121 GWAS included data from 765,348 individuals (567,460 EA, 165,726 of East Asian ancestry, 13,842 African Americans, 13,359 of South Asian ancestry, and 4,961 Hispanics, **Supplementary Table 1**). The median of the study-specific mean eGFR was 89 ml/min/1.73m² (1st (Q1) and 3rd (Q3) quartiles: 81, 94), the median age was 54 years, and 50% were female. GWAS of eGFR were based on genotypes imputed using reference panels from the Haplotype Reference Consortium (HRC)[23] or the 1000 Genomes Project[24] (Methods, **Supplementary Table 2**). Following study-specific variant filtering and quality control procedures, fixed-effects inverse-variance weighted meta-analysis was conducted (Methods). There was no apparent evidence of stratification (LD Score regression intercept =1.04; λ_{GC}=1.05). After variant filtering, 8,221,591 single nucleotide polymorphisms (SNPs) were used for analysis (Methods).

We identified 308 loci containing at least one SNP associated with eGFR at genome-wide significance ($P<5\times10^{-8}$). Loci were defined as ±500 kb around the SNP with the lowest p-value (index SNP); the extended MHC region was considered as a single locus (Methods). Of these loci, 200 were novel and 108 were considered known because they contained an index SNP reported by previous GWAS of eGFR (**Figure 1**). Association statistics for all 308 index SNPs are shown in **Supplementary Table 3** and the corresponding regional association plots in **Supplementary Figure 2**. The minor alleles across index SNPs showed both decreasing and increasing effects on eGFR (**Figure 1**, inset). Taken together, the 308 index SNPs explained 7.1% of the eGFR variance (Methods), nearly doubling the estimate from a recent eGFR GWAS meta-analysis.[9] These SNPs explained 19.6% of the genetic heritability of eGFR (h^2=39%, 95% credible interval: 32%, 47%) estimated in a large participating general-population-based pedigree-study (**Supplementary Figure 3A**; Methods).

Most of the 308 index SNPs showed homogeneous effects across studies (median I^2 =5%, Q1, Q3: 0, 13%; **Supplementary Table 3**; **Fig 2A**). Only one index SNP had I^2 >50% (*UMOD-PDILT* locus, I^2=60%) as described previously;[10,13] heterogeneity is suspected to be age-related.[25] We then investigated the heterogeneity of genetic effects that was correlated with ancestry using meta-regression[26] (Methods) and identified three index SNPs with significant ancestry-related heterogeneity at the *LINC01362*, *GATM*, and *PSD4* loci (ancestry heterogeneity p-value (p-anc-het)

<0.05/308; **Figure 2A**, **Supplementary Table 3**). The index SNP at *UMOD-PDILT* did not show evidence for ancestry-related heterogeneity (p-anc-het=0.59). These results suggest similar effects across ancestries for the vast majority of the identified SNPs. Ancestry-specific results for all 308 index SNPs are reported in **Supplementary Table 4**.

Generalizability of eGFR-associated SNPs and meta-analysis of >1 million individuals

Next, we assessed whether the trans-ethnic findings were generalizable to other, independent trans-ethnic samples, using data from a GWAS meta-analysis of eGFR performed among 280,722 participants of the Million Veteran Program (MVP) from US Veterans' Administration facilities.[27] The MVP sample consisted of 20.4% African American and 79.6% EA participants (mean age 64 years, 7.4% females; Methods). Despite differences in source population, ancestry and sex composition, and analytical strategies between CKDGen and MVP, associations at 262 of the 305 SNPs replicated (1-sided replication $P<0.05$ and $P<5\times10^{-8}$ in the meta-analysis of CKDGen and MVP; **Figure 1**; **Supplementary Table 3**; Methods). Of note, 10 SNPs that were not replicated were previously reported in other GWAS of eGFR, which generally included a replication stage. They may therefore represent true associations that did not replicate in the MVP because of differences between the two studies. Generalizability of the 305 SNPs was further supported by the almost perfect direction consistency of the eGFR effect estimates in CKDGen and MVP (302/305 SNPs, 99%; **Figure 2B**), and the very strong correlation of the genetic effects (Pearson's correlation coefficient ρ=0.92, 95% confidence interval (CI): 0.90, 0.94; **Figure 2B**). Seventy-four SNPs were genome-wide significant in MVP alone (**Supplementary Table 3**). The meta-analysis of CKDGen and MVP comprising a total of 1,038,119 individuals showed genome-wide significance for 94% of the 305 SNPs (**Supplementary Table 3**), with the lowest p-value being observed for SNP rs77924615 at *UMOD-PDILT* ($P=3\times10^{-259}$). Together, these results support the robustness of our findings even in a setting of a smaller independent study with differences in ancestral and sex composition, trait transformation, and modeling (Methods).

Clinical and epidemiological landscape: association of eGFR index SNPs with blood urea nitrogen (BUN) and CKD

To evaluate whether the eGFR index SNPs pertained to kidney function rather than creatinine metabolism, we assessed their relation to BUN, an alternative marker of kidney function that is inversely correlated with eGFR. We performed fixed-effect meta-analysis of GWAS of BUN combining 416,178 samples from 65 studies (**Supplementary Table 1**) across 5 ancestries, using the same workflow as for eGFR.

Overall, the GWAS meta-analysis of BUN showed no evidence of stratification (λ_{GC}=1.03; LD Score regression intercept =0.98) and yielded 111 genome-wide significant loci (15 known, 96 novel, **Supplementary Figure 4**, **Supplementary Table 6**). The 262 replicated eGFR index SNPs showed an inverse correlation with BUN (ρ=-0.65, 95%CI: -0.72, -0.58). Evidence of their relevance to kidney function was classified as *strong* for 148 eGFR index SNPs that showed inverse, significant (one-sided P<0.05) associations with BUN; *inconclusive* for 102 eGFR index SNPs that were not significantly associated with BUN (P≥0.05); and *unlikely* for 15 eGFR index SNPs showing concordant, significant (one-sided P<0.05) association with BUN (**Figure 2C, Supplementary Table 5**). This comparative analysis of complementary markers of kidney function supports that signals at most of the eGFR-associated index SNPs likely reflect relevance for kidney function rather than marker metabolism.

Next, we investigated the odds ratios (OR) for CKD, defined as eGFR <60 ml/min/1.73m^2, for all eGFR index SNPs. We performed a GWAS meta-analysis of CKD in 625,219 individuals (64,164 CKD cases) from the same studies used for eGFR analysis (Methods). Of the 262 replicated eGFR index SNPs, 222 were significantly associated with CKD (1-sided P<0.05, **Figure 1, inset**), including 128 with strong evidence of kidney-function relevance as classified by their association with BUN (**Supplementary Table 5**). Twenty-three loci were associated with CKD at genome-wide significance level. Of them, two haven't been found in previous GWAS of eGFR or CKD, and 17 showed strong support for kidney function relevance, at *SDCCAG8*, *PSD4*, *HOXD8*, *FGF5*, *DAB2*, *SLC34A1*, *UNCX*, *PRKAG2*, *LARP4B*, *DCDC1*, *WDR72*,

UMOD-PDILT, *MYO19*, *AQP4*, *NFATC1*, and *NRIP1* (**Supplementary Table 5**). The largest effects on CKD were observed for rs77924615 at *UMOD-PDILT* (OR=0.81, 95%CI: 0.80; 0.83), rs187355703 at *HOXD8* (OR=0.82, 95%CI: 0.77; 0.87), and rs10254101 at *PRKAG2* (OR=1.11, 95%CI: 1.09; 1.11).

Lastly, we carried out a genetic risk score (GRS) analysis in a large, independent dataset of 452,264 participants of the UK Biobank to assess the combined effect of identified eGFR index SNPs on clinically diagnosed CKD and CKD related outcomes including hypertension, defined by ICD-10 codes (Methods). Each standard deviation higher GRS, constructed based on effect-size weighted alleles associated with lower eGFR across the index SNPs, was significantly associated with higher odds of chronic renal failure (OR 1.08, 95%CI: 1.07, 1.09; P=8.7x10^{-59}). Moreover, the GRS also showed positive and significant associations with glomerular diseases, acute renal failure, and hypertensive diseases, but not with asthma or schizophrenia that were included as negative controls (**Figure 2D**, **Supplementary Figure 5**). There was a significant protective association with urolithiasis, which may reflect a reduced ability to concentrate urine at lower eGFR.

Genetic correlations of eGFR and BUN with other complex traits and diseases

Lower eGFR is associated with numerous cardio-metabolic risk markers and diseases.[4] We therefore assessed the evidence for a shared genetic basis or co-regulation by evaluating genetic correlations (r_g) of the trans-ethnic GWAS meta-analysis results with those from 749 other complex traits and diseases available through LD Hub using LD score regression (Methods).[28] For eGFR, we observed 37 significant genetic correlations (P<6.7×10^{-5} = 0.05/749, **Supplementary Figure 6**; **Supplementary Table 8**). Other than with serum creatinine, the largest negative genetic correlations were observed for serum citrate (r_g=-0.27) and urate (r_g=-0.23), followed by anthropometric traits including lean mass and physical fitness (e.g., r_g=-0.20 with left hand grip strength). While the inverse genetic correlation with muscle mass-related traits likely reflects higher creatinine generation leading to lower creatinine-based eGFR, the genetic correlations with blood concentrations of the metabolites citrate and

urate likely reflect reduced filtration function, as does the positive genetic correlation with GFR estimated from cystatin C (r_g=0.53). No significant genetic correlations were observed with cardiovascular traits and diseases. Sensitivity analyses using only EA-specific summary statistics delivered very similar results (data not shown). For BUN, a very similar pattern of genetic correlations with the 748 other complex traits and diseases was observed as for eGFR (**Supplementary Table 8**). Importantly, the strongest genome-wide genetic correlation across all traits was observed with CKD (r_g=0.47).

In summary, significant genetic correlations with eGFR reflect the two biological components that govern serum creatinine concentrations: its excretion via the kidney and its generation in muscle. The high proportion of traits and diseases that show significant genetic correlations with both eGFR and BUN support the use of BUN as a comparative biomarker to prioritize loci most likely to reflect effects on kidney function rather than marker metabolism.

Functional enrichment and pathway analyses

To identify previously unknown molecular mechanisms and tissues of importance for kidney function, we assessed the enrichment of the eGFR and BUN genetic associations using tissue-specific gene expression, regulatory annotations, and gene sets and pathways (Methods).

First, we used all eGFR-associated SNPs with $P<5\times10^{-8}$ to explore enriched pathways, tissues and cell types based on gene expression data using DEPICT.[29] We identified 16 significantly enriched physiological systems, cell types and tissues highlighting several aspects of kidney function, physiology and disease. The strongest enrichment was observed for urogenital and renal physiological systems and tissues (kidney, kidney cortex, and urinary tract; false discovery rate (FDR) <0.05; **Supplementary Figure 7A and B**), which is proof-of-concept of the kidney as the primary target organ. We additionally found significant enrichment for mucous membrane, respiratory mucosa, nasal mucosa, and nose (enrichment p-values from 3.1×10^{-4} to 1.2×10^{-3}), possibly reflecting epithelial cell processes including transport

mechanisms that are shared with the kidney. Pathway and gene set enrichment analysis identified three highly correlated and strongly associated meta gene sets ($P<1\times10^{-6}$, FDR<0.05), including some relevant to kidney such as polyuria, dilated renal tubules, and expanded mesangial matrix, as well as signaling and transcription, and energy metabolism (**Supplementary Figure 7C**). Tissue and cell-type enrichment analysis of BUN-associated SNPs with $P<5\times10^{-8}$ highlighted a very similar pattern to the one observed for eGFR: the strongest enrichment was observed for urogenital and renal physiological systems and tissues (**Supplementary Figure 8**), and significant enrichment was also observed for mucous membrane, respiratory mucosa, nasal mucosa, nose, epithelial cells and the epithelium. In contrast, we did not observe enrichment for muscle tissues as was observed for eGFR, further supporting the use of BUN to prioritize loci most likely to be related to kidney function rather than creatinine metabolism.

Second, we used the genome-wide eGFR and BUN summary statistics to identify cell-type groups with evidence for enriched heritability based on data from diverse, cell-type specific functional genomic elements using stratified LD Score regression.[30] Across 10 evaluated cell type groups, the strongest and most significant enrichment for eGFR was observed for kidney (13.2-fold), followed by liver (7.3-fold) and adrenal/pancreas (5.7-fold enrichment; **Supplementary Table 9**). Kidney cell types were also the most enriched cell type group for BUN (significant 11.5-fold enrichment; **Supplementary Table 9**). These analyses based on regulatory marks confirm the importance of the kidney as a target organ.

Lastly, we took a complementary approach to assess enrichment of eGFR-associated variants in genes resulting in kidney phenotypes in genetically manipulated mice.[31] We selected all genes from the Mouse Genome Informatics database that - when genetically manipulated - cause abnormal GFR (n=24), abnormal kidney physiology (n=453), or abnormal kidney morphology (n=764). Human orthologs for these genes were interrogated in the eGFR summary statistics for the presence of significant associations (Methods). We identified associations in 10 genes causing abnormal GFR in mice (enrichment p-value=8.9×10^{-4}), 55 causing abnormal kidney physiology (enrichment p-value=1.1×10^{-4}) and 96 causing abnormal kidney morphology

(enrichment p-value=1.8×10^{-5}; **Figure 3**, Methods). Of these, 25 genes were novel, i.e. these genes had not previously been shown to contain SNPs associated with eGFR at genome-wide significance or map near known loci (**Supplementary Table 10**). These genes therefore represent additional novel eGFR-associated candidates in humans. The existing mouse models may pave the way for experimental confirmation of these findings.

Statistical fine-mapping and second signal analysis in EA individuals

To effectively fine-map loci using summary statistics, the LD reference panel needs to be ancestry-matched to the GWAS population and the sample size of the LD reference panel should scale with that of the GWAS.[32] Accordingly, conditional and fine-mapping analyses were only carried out among EA participants, for whom data to construct a large enough LD reference panel is publicly available; 15,000 EA individuals from the UK Biobank were randomly selected for this purpose (Methods).

The meta-analysis of the 85 EA GWAS studies identified 256 genome-wide significant loci (**Supplementary Table 11**). Combining neighboring loci whose index SNPs were correlated ($r^2\geq0.2$) resulted in 212 non-overlapping regions. Among these regions, 277 independent genome-wide significant SNPs were identified (Methods, **Supplementary Table 12**), that explained 6.8% of the eGFR variance and 23.2% of eGFR genetic heritability in EA (**Supplementary Figure 3B**). For each of the 277 independent variants, we computed a 99% credible set, that is, a set which contains the SNP driving the association with 99% probability.[33] The median number of SNPs per credible set was 30 (Q1, Q3: 7, 74). Twenty credible sets contained only a single SNP (at *EDEM3*, *CACNA1S*, *HOXD11*, *CPS1*, *DAB2*, *SLC34A1*, *LINC01512*, *LARP4B*, *DCDC1*, *SLC25A45*, *SLC6A13*, *GATM*, *CGNL1*, *CYP1A1*, *NRG4*, *RPL3L*, *UMOD-PDILT, SLC47A1*), and two independent sets with a single SNP each at *BCL2L14* (**Supplementary Table 12; Figure 4**), and 58 sets contained ≤5 SNPs (small credible set).

To ensure that loci we prioritize for further functional studies represent robust findings with likely relevance for kidney function, selected credible sets were restricted

to those based on SNPs that replicated in the MVP and which showed tier 1 or 2 relevance for kidney function. Associations at the 256 EA-specific index SNPs were therefore tested for replication in the EA subsample of the MVP using the same criteria as in the trans-ethnic analysis, and confirmed association at 228 loci (Methods, **Supplementary Table 11**). GWAS meta-analysis of BUN was performed in 243,029 EA individuals, and BUN association results at 15 of 228 replicated eGFR index SNPs indicated that these were *unlikely* to be related to kidney function (**Supplementary Table 11**). This left 253 credible sets that contain genes and SNPs that can be prioritized for further study (**Supplementary Table 12**).

To systematically examine the characteristics of the SNPs in the credible sets, SNPs were annotated with respect to their functional consequence and regulatory potential. Missense SNPs with a posterior probability of >50% of driving the association and/or mapping into a small credible set are of particular interest because they directly implicate the affected gene. Such missense SNPs were identified in 11 genes (*SLC47A1*, *RPL3L*, *SLC25A45*, *CACNA1S*, *EDEM3*, *CPS1*, *KLHDC7A*, *PPM1J*, *CERS2*, *C9,* and *SLC22A2;* **Supplementary Table 13**). Most of these variants had CADD scores >15 (**Figure 4A**, **Table 1**), a cutoff used to indicate deleteriousness.[34] As summarized in **Table 1**, several of the identified genes are plausible biological candidates. For example, the missense p.Ala465Val SNP in *SLC47A1* (posterior probability >99%) encodes an amino acid change of the encoded multidrug and toxin extrusion protein (MATE1). This transport protein is responsible for the secretion of cationic drugs, toxins and internal metabolites including creatinine across brush border membranes including kidney proximal tubules. The fact that MATE1 knockout mice have higher blood levels of both creatinine and BUN[35] argues against a sole effect on creatinine transport. Altered ability to excrete toxic compounds via kidney tubular cells may also be the molecular mechanisms underlying the association signal at another fine-mapped missense SNP, p.Ser270Ala in *SLC22A2* (**Table 1**).

Emerging experimental evidence provides molecular mechanisms by which regulatory variants identified from GWAS exert their effects.[36] To evaluate whether small credible set SNPs may have regulatory potential in the kidney, we annotated them to regions of open chromatin identified from primary cultures of human tubular and

glomerular cells (GEO accession number GSE115961), as well as from publicly available kidney cells types (ENCODE and Roadmaps Projects; Methods). We identified 63 SNPs in 39 credible sets that mapped into one of these annotations and may thus represent causal regulatory variants (**Supplementary Table 13**). A finding of particular interest was intronic rs77924615 in *PDILT*. This SNP had a posterior probability of >99% of driving the association signal at the well-established *UMOD* locus. It mapped into open chromatin in all evaluated resources (native kidney cells, ENCODE and Roadmap kidney cell types), implicating rs77924615 as a candidate causal regulatory variant (**Figure 4B**) associated with differential expression of uromodulin, the product of the neighboring *UMOD* gene, but not with the *PDILT* transcript in any tissue.

Gene prioritization via gene expression co-localization analysis

A complementary approach to highlight target genes in associated loci is to systematically evaluate co-localization of the genetic associations with phenotype and gene expression (expression quantitative trait locus, eQTL) in *cis*. We performed co-localization analyses for each eGFR-associated locus with gene expression across 46 tissues including kidney glomerular and tubulo-interstitial compartments (Methods). A high posterior probability of >80% for co-localization of eGFR and eQTL signals in at least one kidney tissue was observed for 20 transcripts (**Figure 5**), pointing towards a shared underlying SNP and implicating the gene encoding for the co-localized transcript as the locus' effector gene(s).

Novel insights emerged on several levels: first, *UMOD* is a well-established causal gene for CKD in the associated GWAS locus and can therefore serve as a validation of the workflow. In the tubulo-interstitial kidney compartment, we observed evidence for a shared underlying variant associated with higher *UMOD* gene expression and lower eGFR (**Figure 5**). This is consistent with previous GWAS of urinary uromodulin concentrations, in which alleles associated with lower eGFR at *UMOD*[15] were associated with higher urinary uromodulin concentrations.[37] The lead SNP at this locus was rs77924615, highlighted above as the candidate causal regulatory variant mapping into the intron of *PDILT*, the gene upstream of *UMOD*. The association with

differential *UMOD* but not *PDILT* gene expression supports *UMOD* as the causal gene at this locus and rs77924615 as a regulatory SNP. It also illustrates the value of studying gene expression in diverse tissues, as this kidney-specific co-localization would have been missed had target tissue not been studied.

Second, novel, biologically plausible candidates emerged. For example, our results suggest that *KNG1* and *FGF5* are the effector genes in the respective eGFR-associated loci (**Figure 5**, **Supplementary Table 14**). *KNG1* encodes for high-molecular weight kininogen, which among other functions releases bradykinin. Bradykinin in turn influences blood pressure, natriuresis and diuresis, and can be linked to kidney function via the connection to the renin-angiotensin-aldosterone system.[38] Independent variants (r^2<0.1) at the *KNG1* locus have been identified in GWAS of plasma renin and aldosterone concentrations,[39] and a perfect proxy of the *KNG1* index SNP was reported in GWAS of adiponectin concentrations.[40] *FGF5* encodes for Fibroblast Growth Factor 5, and the index SNPs for eGFR or highly correlated SNPs (r^2>0.9) have been identified in multiple GWAS of blood pressure, atrial fibrillation, coronary artery disease, hematocrit and multiple kidney-function related traits (**Supplementary Table 14**). The eGFR index SNP, rs1458038, has a posterior probability of >50% of driving the association signal with eGFR and has a CADD score of 14.8 (**Supplementary Table 14**, supporting its regulatory potential on the expression levels of *FGF5* transcript levels, which co-localized with the eGFR signal only in tubule-interstitial kidney portions (**Figure 5**). Index SNPs at both *KNG1* and *FGF5* are likely to be related to kidney function as they also show associations with BUN and CKD.

Examination of gene expression across multiple tissues types revealed some of the kidney-colocalized genes showed directionally consistent transcript levels across all tissues (e.g. *METTL10*), while other kidney-colocalized genes were bidirectional, with higher transcript levels in some tissues, but lower in others (e.g. *SH3YL1*) (**Figure 5**). These observations were also reflected broadly across all genes with evidence of colocalization in any tissue (**Supplementary Figure 8**), and highlight tissue-shared and tissue-specific signals.[41,42]

Trans-eQTL annotation of the index SNPs was only performed using whole blood and peripheral blood mononuclear cells, for which eQTL datasets with large sample size were available (Methods). Based on the analysis of 5 non-overlapping EA genome-wide eQTL studies (sample size range 1469 - 6645, **Supplementary Table 15**), we identified a reproducible link of rs17696736 (12q24.12) with both the calcium-binding protein gene *S100A10* (1q21.3) and *STAT1* (2q32.2). *S100A10* encodes a subunit of annexin A2, which co-localizes with *PLA2R* at the cell surface and in extracellular vesicles from podocytes.[43] Inhibition of *STAT1* has been reported to protect from glomerular mesangial cell senescence[44] and to ameliorate renal oxidative stress[45] (**Supplementary Table 16**).

Co-localization with protein levels supports UMOD as a target gene

The *UMOD* locus is of particular clinical interest for CKD research:[20] rare *UMOD* mutations cause autosomal-dominant tubulo-interstitial kidney disease[46] and common variants at *UMOD* give rise to the strongest signal in GWAS of eGFR and CKD.[15] We therefore investigated this locus in further detail: conditional analyses based on the EA-specific summary statistics indicated the presence of two independent variants at this locus (**Figure 6A**), with rs77924615 mapping into upstream *PDILT*, and rs34882080 mapping into an intron of *UMOD*. Association results for the urinary uromodulin-to-creatinine ratio (UUCR) in one of the participating cohorts gave a similar appearance (**Figure 6B**) with rs34262842 (r^2 with rs34882080 0.93) as the lead variant. Co-localization of the associations with eGFR and with UUCR was evaluated separately for the two independent signals, the one with lead variants in *UMOD* (**Figure 6C**) and the one represented by rs77924615 in *PDILT* (**Figure 6D**). In both regions, there was a high probability of a shared underlying variant (posterior probabilities of 0.97 and 0.96, respectively), further supporting rs77924615 as a causal regulatory variant and *UMOD* as its effector gene.

Discussion

This trans-ethnic study represents a 5-fold increase in sample size compared to previous GWAS meta-analyses of eGFR and identified 308 eGFR-associated loci, of which 200 are reported here for the first time. The index SNPs at these loci explain almost twice as much eGFR variance as previously reported.[9,10,22] By using complementary kidney function traits, we highlight loci that most likely reflect the kidney's filtration function and provide a comprehensive annotation resource. Our enrichment analyses confirm the kidney as the main target tissue of the detected SNPs, and co-localization with gene expression in kidney prioritize 18 potential target genes for follow-up. Conditional analyses followed by statistical fine-mapping and annotation implicated a single potentially causal SNP in 20 independent loci, and identified 10 missense SNPs directly implicating *RPL3L*, *CACNA1S*, *CERS2* and *C9* as effector genes. The increase in resolution from locus to single potentially causal variant with its effector gene and target tissue, as illustrated by in depth analysis of the *UMOD* locus, represents a critical advance in the field and is a prerequisite for translational research.

Most previous meta-analyses of GWAS of eGFR were limited to a single ancestry group[8] and did not prioritize causal variants or effector genes in associated loci. While being underpowered to uncover novel loci, one previous trans-ethnic study employed statistical fine-mapping and resolved one signal to a single variant,[19] rs77924615 at *UMOD-PDILT*, also identified in our study. At this locus, we further characterized the relationship between the causal variant at *PDILT*, *UMOD* expression in the target tissue, and UMOD protein levels. This represents a significant advancement over the course of 10 years of eGFR GWAS[15] and highlights the utility of the generated resources.

Our complementary multi-tissue approaches including enrichment analyses based on gene expression, regulatory annotations, and gene sets and pathways highlight the kidney as the most important target organ. However, relatively few kidney-specific experimental datasets are publicly available as compared to other organs and tissues. For example, the kidney is not well represented in the GTEx Project and not included in its tissue-specific eQTL datasets,[41] which emphasizes the value of open

access resources and in depth characterization of uncommon tissues and cell types. We were able to specifically investigate the kidney by using a recently published eQTL dataset from glomerular and tubulo-interstitial portions of micro-dissected human kidney biopsies,[47] kidney-specific regulatory information from the ENCODE and Roadmap resources, and by obtaining regulatory information from primary cultures of human glomerular and tubulo-interstitial cells.

Functional follow-up studies of potentially causal variants will benefit from prioritized loci that show clear evidence supporting one or a few variants driving the association signal. Our statistical fine-mapping workflow allowed us to prioritize such variants at single-variant resolution for 20 loci, and down to a set of ≤5 SNPs for 38 additional loci. For example, the OCT2 protein encoded by the prioritized *SLC22A2* gene is known to transport several cationic drugs such as metoprolol, cisplatin, metformin and cimetidine across the basolateral membrane of renal tubular cells.[48] The p.Ser270Ala SNP prioritized by our workflow is a known pharmacogenomic variant that alters the transport of these drugs and their side-effects, such as cisplatin-induced nephrotoxicity.[49] Some of these drugs are commonly prescribed to CKD patients and may be of relevance given their already reduced eGFR. Along the same lines, the prioritized p.Ala465Val SNP in *SLC47A1* that encodes the transporter MATE1 protein may affect the ability to secrete drugs and other toxins from proximal tubular cells into the urine[50] and hence alter CKD risk.

Strengths of this project include the large trans-ethnic sample size with dense genotype imputation, a standardized and automated phenotype generation and quality control workflow, the use of an alternative biomarker of kidney function to prioritize associated loci with respect to their kidney function relevance, associations of a genetic risk score for low eGFR with CKD and related diagnoses in a large independent population, and advanced and comprehensive downstream bioinformatics analyses to prioritize causal genes and SNPs across tissues. Some limitations warrant mention. Non-European populations were still underrepresented in our study, emphasizing the need study more diverse populations[51] and highlighting the potential of future trans-ethnic efforts with trans-ethnic fine-mapping analyses once larger reference panels to estimate population-specific LD become available. Several SNPs had small effective

sample sizes in some of the subpopulations, which might have affected the ability to assess between-ancestry heterogeneity and potentially resulted in an underestimation of true heterogeneity. We used GFR estimated from serum creatinine, as done in clinical practice and observational studies, because direct measurement of kidney function is invasive, time-consuming, and burdensome. We carefully calibrated creatinine across studies, used state-of-the-art estimating equations, and distributed a centrally generated and automated script for GFR phenotype computation that all participating studies used. We also evaluated genetic associations with a complementary marker of kidney function, BUN. In addition, the generated genome-wide BUN summary statistics represent a useful resource for other studies in the field that evaluate only one kidney function biomarker, typically eGFR, or for researchers interested in organ-specific functions. To favor the identification of signals and variants that are broadly representative and generalizable, our analysis focused on SNPs that were present in the majority of the participating studies. This choice might have limited our ability to uncover novel or to fine-map low-frequency or population-specific variants, which represents a complementary avenue of research. Moreover, even with well-powered statistical fine-mapping approaches, prioritized variants estimated as underlying the association signals need to be confirmed as causal variants in experimental studies. Although co-localization with gene expression can help prioritize effector genes, these associations are based on measures from a single time point and hence cannot answer whether changes in gene expression precede changes in kidney function or occur as a consequence.

In summary, we identified and characterized 308 loci associated with eGFR and prioritized potential effector genes, driver variants and target tissues. These findings will help direct functional studies and advance the understanding of kidney function biology, a prerequisite to develop novel therapies to reduce the burden of CKD.

Online Methods

Overview

We set up a collaborative meta-analysis based on a distributive data model and quality control (QC) procedures. To maximize the level of standardization of generated phenotypes across studies, an analysis plan and a command line script (https://github.com/genepi-freiburg/ckdgen-pheno) were created centrally and provided to all participating studies, which were mostly population-based (**Supplementary Table 1**). Instructions for data processing, analysis and troubleshooting were distributed to all studies via a Wiki system (https://ckdgen.eurac.edu/mediawiki/index.php/CKDGen_Round_4_EPACTS_analysis_plan). Automatically generated summary files were uploaded centrally for phenotype quality approval of the generated phenotypes. GWAS were then run within each study and uploaded centrally. GWAS QC was performed using GWAtoolbox[52] and custom scripts to assess ancestry-matching allele frequencies and fix variant positions. All studies had their own research protocols approved by the respective local ethics committees. All participants in all studies provided written informed consent.

Phenotype definition

Each study measured serum creatinine and blood urea nitrogen (BUN) as described in **Supplementary Table 1**. When measured with a Jaffé assay before 2009, serum creatinine values were calibrated by multiplying by 0.95.[53] In studies of >18 year-old adults, eGFR was estimated with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation,[54] using the R software package 'nephro'.[55] In studies of ≤18 year-old subjects, eGFR was estimated with the Schwartz formula[56]. eGFR values were winsorized at 15 and 200 ml/min/1.73 m^2. CKD was defined as eGFR <60 ml/ min/1.73 m^2. BUN was derived for studies that reported blood urea measurements by multiplication by 2.8, and units were aligned to mg/dl across cohorts. All steps occurred in the central phenotyping script.

Genotyping and genotype imputation

Genotype imputation was conducted based on the Haplotype Reference Consortium (HRC) version 1.1 or the 1000 Genomes Project phase 3 version 5 (1000Gp3v5) ALL or phase 1 version 3 (1000Gp1v3) ALL panels. Imputed variants were coded as allelic dosages accompanied by the corresponding imputation quality (IQ) scores (ImputeV2 info score, MACH/minimac RSQ, or as applicable), and annotated on the NCBI b37 (hg19) reference build. Study-specific genotyping arrays as well as haplotype phasing and genotype imputation methods are described in **Supplementary Table 2**.

Genome-wide association studies (GWAS)

In each study, eGFR residuals were derived from sex- and age-adjusted linear regression models fitted to log(eGFR) and BUN using the centrally distributed script. In the subsequent GWAS, residuals were regressed on SNP dosage levels, assuming an additive genetic model. Study-specific features, such as study site, genetic principal components (PCs), or relatedness, were accounted for in the study-specific models as appropriate (**Supplementary Table 2**). Logistic regression models were fitted for CKD.

Trans-ethnic GWAS meta-analysis

For eGFR, 121 GWAS summary statistics files were contributed across studies. After QC, the total samples size was 765,348 (567,460 individuals of European ancestry [EA], 165,726 of East Asian ancestry, 13,842 African Americans, 13,359 of South Asian ancestry, and 4961 Hispanics; **Supplementary Table 1**). For CKD, 60 GWAS summary files were contributed, totaling a post-QC samples size of 625,219 (64,164 cases). For BUN, 65 GWAS summary files were contributed, totaling a post-QC samples size of 416,178 (see **Supplementary Table 1** for ancestry-specific details of CKD and BUN).

Before meta-analysis, study-specific GWAS files were filtered to retain only variants with IQ score >0.6 and minor allele count (MAC) >10. Within study, we estimated the genomic control (GC) factor λ_{GC} and applied GC correction when λ_{GC} was >1. Fixed effects inverse-variance weighted meta-analysis was performed using

METAL,[57] which was adapted to increase effects and standard errors precision (seven decimal places instead of four).

After meta-analysis of 43,994,957 SNPs, we retained only variants that were present in ≥50% of the GWAS data files and had a total MAC of ≥400. Across ancestries, this yielded 8,221,591 variants for eGFR (8,834,748 in EA), 8,176,554 variants for BUN (8,358,347 in EA), and 9,585,923 variants for CKD. Post-meta-analysis GC correction was not applied because the LD Score regression intercept was close to 1 in all analyses of eGFR, BUN, and CKD.[58] The genome-wide significance level was set at 5×10^{-8}. Between-study heterogeneity was assessed using the I^2 statistic.[59] For CKD, variants with an I^2 ≥95% were removed to avoid exaggerated influence of single large studies. Variants were assigned to loci by selecting the SNP with the lowest p-value genome-wide as the index SNP, defining the corresponding locus as the 1 Mb-segment centered on the index SNP, and repeating the procedure until no further genome-wide significant SNPs remained. A locus was considered novel if it did not contain any variant identified by previous GWAS of eGFR.

Meta-regression analysis of trans-ethnic GWAS

For eGFR, we evaluated heterogeneity attributable to ancestry using quality-controlled study-specific GWAS files and the software Meta-Regression of Multi-Ethnic Genetic Association (MR-MEGA v0.1.2.[26]). Meta-regression models included three axes of genetic variation. GC correction was applied to the meta-regression results. For the 308 genome-wide significant index SNPs from the trans-ethnic GWAS meta-analysis, we tested ancestry-related heterogeneity of allelic effects at a significance level of 0.05/308 = 1.6×10^{-4} (indicating the corresponding p-value as p-anc-het).

Proportion of phenotypic variance explained and genetic heritability analysis

The proportion of phenotypic variance explained by the index SNPs was estimated as $\beta^2\left(\frac{2p(1-p)}{var}\right)$, with β being the SNP effect, p the effect allele frequency, and var the variance of the sex- and age-adjusted log(eGFR) residuals (assumed as 0.016 based

on data from 11,827 EA participants of the population-based ARIC study).[9] The proportion of variance explained by independent genome-wide significant index SNPs was estimated using the GCTA COJO Slct analysis (see fine-mapping section below).

Genetic heritability of age- and sex-adjusted log(eGFR) was estimated using the R package 'MCMCglmm'[60] in the Cooperative Health Research In South Tyrol (CHRIS) study,[61] a participating pedigree-based study of EA individuals (186 up-to-5 generation pedigrees, totaling 4373 subjects).[62] We fitted two models, with and without the inclusion of the identified index variants (304/308 and 277/277 from the transethnic and EA analyses, respectively), running 1,000,000 MCMC iterations (*burn in* = 500,000) based on previously described settings.[62]

Comparison with and replication of results in the Million Veteran's Program (MVP)

To understand the robustness and generalizability of the eGFR-associated SNPs identified in the CKDGen Consortium, we interrogated the effect estimates of the 308 trans-ethnic index SNPs in a GWAS from an independent, large, trans-ethnic study, the Million Veteran Program (MVP).[27] Briefly, the MVP study participants were recruited across 63 U.S. Veteran's Administration (VA) medical facilities. Written informed consent was obtained and all documents and protocols were approved by the VA Central Institutional Review Board. DNA was genotyped using a customized Affymetrix Axiom Biobank Array chip with additional content added to provide coverage of African and Hispanic haplotypes, as well as markers for common diseases in the VA population. After QC, genotype were pre-phased using EAGLE version 2[63] and imputed based on the 1000Gp3v5 reference panel using minimac3.[64] Genotype PCs were estimated using FlashPCA.[65] Serum creatinine was assessed up to one year prior to MVP enrollment using isotope dilution mass spectrometry. eGFR was estimated using the CKD-EPI equation[54] after excluding subjects on dialysis, transplant patients, amputees, individuals on HIV medications, and those with creatinine values of <0.4 mg/dl. Diabetes was defined as use of anti-diabetic medications or by assignment of an International Classification of Diseases 9 (ICD-9) code for diabetes during the baseline period. Hypertension was defined as having an ICD-9 code for hypertension, being on antihypertensive drug or having ≥2 measures of systolic or diastolic blood pressure >140 mmHg or >90 mmHg, respectively.

GWAS of eGFR on SNP dosage levels were performed by fitting linear regression models adjusted for age at creatinine measurement, age^2, sex, body mass index, and the first 10 genetic PCs, using SNPTEST version 2.5.4-beta.[66] All GWAS were stratified by self-reported ethnicity (79.6% White non-Hispanic and 20.4% Black non-Hispanic), diabetes, and hypertension status. Results were combined across strata using fixed effects inverse-variance weighted meta-analysis in METAL.[57] This analysis encompassed a total of 280,722 subjects across all strata, of whom 216,518 were non-Hispanic Whites (EA).

Of the 308 trans-ethnic eGFR index variants, 305 variants or their good proxies were available for replication in the MVP GWAS (proxies had to have $r^2 \geq 0.8$ with the index SNP and were selected by maximum r^2 followed by minimum distance in case of ties). Replication testing of the 256 EA-specific index SNPs was restricted to the MVP EA GWAS. CKDGen and MVP meta-analysis results were pooled via sample size weighted meta-analysis of z-scores using METAL.[57] In both the trans-ethnic and EA-specific analyses, replication was defined as a one-sided p-value <0.05 in the MVP and genome-wide significance of the CKDGen and MVP meta-analysis result.

Classification of kidney function relevance of eGFR-associated loci based on BUN

We used genetic associations with BUN to classify replicated index SNPs for eGFR with respect to their potential kidney function relevance into three tiers of evidence: tier 1 ("*strong support for kidney function relevance*") included all eGFR index SNPs with an inverse, significant (one-sided $P<0.05$) association with BUN for a given reference allele, tier 2 ("*inconclusive evidence of kidney function relevance*") included all eGFR index SNPs whose effect on BUN was not significantly different from 0 (one-sided $P>0.05$); and tier 3 ("*kidney function relevance is unlikely*") included all eGFR index SNPs with a concordant, significant (one-sided $P<0.05$) association with BUN for a given reference allele.

Genetic risk score analysis in the UK Biobank dataset

Genetic risk score (GRS) analyses were carried out based on summary statistics from 452,264 participants of the UK Biobank. Genome-wide association results were obtained from GeneAtlas[67] for the following ICD-10 codes: glomerular diseases (N00-N08; number of cases: 2289); acute renal failure (N17; 4913); chronic renal failure (N18; 4905); urolithiasis (N20-N23; 7053); hypertensive diseases (I10-I15; 84,910); and ischemic heart diseases (I20-I25; 33,387). Asthma (J45; 28,628) and schizophrenia (F20; 590) were included as negative controls. Of the 308 eGFR index SNPs from the trans-ethnic GWAS meta-analysis, 302 were available in the UK Biobank dataset. A

genetic risk score was estimated as the combined effect of alleles associated with lower eGFR levels at each of the 302 SNPs. The coefficient and standard error of the risk score is a weighted mean of the per-SNP regression coefficients (matched to the eGFR-decreasing allele) and the combined squared standard errors, respectively, where each is weighted by the effect size of the respective allele from the eGFR CKDGen discovery meta-analysis.[68] The same analysis was repeated restricted to 250 replicated SNPs with kidney function relevance of tier 1 or 2, of which 245 were present in the UK Biobank dataset and were used for GRS calculation (**Supplementary Figure 5**).

Genome-wide genetic correlations with other complex traits and diseases

Genome-wide genetic correlation analysis was carried out to investigate evidence of co-regulation or shared genetic bases between eGFR and BUN and other complex traits and diseases, both known and not known to correlate with eGFR and BUN. We estimated pairwise genetic correlation coefficients (r_g) between the results of our trans-ethnic meta-analyses of eGFR and BUN and each of 749 pre-computed and publicly available GWAS summary statistics of complex traits and diseases available through LD Hub version 1.9.0 using LD Score regression.[28] An overview of the sources of these summary statistics and their corresponding sample sizes is available at http://ldsc.broadinstitute.org. Statistical significance was assessed at the Bonferroni corrected level of 6.7×10^{-5} (=0.05/749).

Functional enrichment: pathway and tissue enrichment analysis

We used DEPICT version 1 release 194 to perform Data-Driven Expression Prioritized Integration for Complex Traits analysis,[29] including pathway/gene-set enrichment and tissue/cell type analyses as described previously.[9,10] All 14,461 gene sets were reconstituted by identifying genes that were transcriptionally co-regulated with other genes in a panel of 77,840 gene expression microarrays,[69] from mouse knock-out studies, and molecular pathways from protein-protein interaction screening. In the tissues and cell type enrichment analysis, we tested whether genes in associated

regions were highly expressed in 209 MeSH annotation categories for 37,427 microarrays on the Affymetrix U133 Plus 2.0 Array platform. For both eGFR and BUN, we included all variants associated with the trait at a p-value of $<5\times10^{-8}$ in the trans-ethnic meta-analysis. Independent variant clumping was performed using Plink 1.9[70] with 500 kb flanking regions and $r^2>0.01$ in the 1000Gp1v3 dataset. After excluding the MHC region, DEPICT was run with 500 repetitions to estimate the FDR and 5000 permutations to compute p-values adjusted for gene length by using 500 null GWAS. All significant gene sets were merged into meta gene sets by running an affinity propagation algorithm[71] implemented in the Python 'scikit-learn' package (http://scikit-learn.org/). The resulting network was visualized using Cytoscape (http://cytoscape.org/).

Enrichment of heritability by cell type group

We used stratified LD Score regression as a complementary method to investigate important tissues and cell types based on the trans-ethnic eGFR and BUN meta-analysis results. Heritability enrichment in 10 cell type groups was assessed using the default options of stratified LD Score regression described previously.[30] The 10 cell type groups were collapsed from 220 cell-type specific regulatory annotations for the four histone marks H3K4me1, H3K4me3, H3K9ac, and H3K27ac. The enrichment of a cell type category was defined as the proportion of SNP heritability in that group divided by the proportion of SNPs in the same cell type group.

Identification of variants in genes causing kidney phenotypes in mice

A nested candidate gene analysis was performed using GenToS[72] to identify additional genetic associations that were missing conventional genome-wide statistical significance. Candidate genes causing kidney phenotypes in mice upon manipulation were selected using the comprehensive Mouse Genome Informatics (MGI) phenotype ontology in September 2017 (abnormal renal glomerular filtration rate [MP:0002847]; abnormal kidney morphology [MP:0002135]; abnormal kidney physiology

[MP:0002136]). The human orthologs of these genes were obtained using the Human-Mouse: Disease Connection webtool (http://www.informatics.jax.org/humanDisease.html). Genes with no human orthologs were removed. Statistical significance was defined as a Bonferroni correction of a type I error level of 0.05 for the number of independent common SNPs across all genes in each of the three candidate gene lists plus their flanking regions, based on an ancestry-matched reference population. In a next step, the GWAS meta-analysis summary statistics for eGFR were queried for significantly associated SNPs mapping into the selected candidate genes. Enrichment of significant genetic associations in genes within each candidate list was computed from the complementary cumulative binomial distribution.[72] GenToS was used with default parameters on each of the three candidate gene lists, using the 1000 Genomes phase 3 release 2 ALL dataset as reference.

Identification of independent variants in the EA meta-analysis

To identify additional, independent eGFR-associated variants within the identified loci, approximate conditional analyses were carried out that incorporated LD information from an ancestry-matched reference population. We used the genome-wide eGFR summary statistics from the EA meta-analysis as input, because an LD reference sample scaled to the size of our meta-analysis was only available for EA individuals.[32] We randomly selected 15,000 participants from the UK Biobank dataset (UKBB; dataset ID 8974). Individuals who withdrew consent and those not meeting data cleaning requirements were excluded, keeping only those who passed sex-consistency check, had ≥95% call rate, and did not represent outliers with respect to SNP heterozygosity. For each pair of individuals, the proportion of variants shared identical-by-descent (IBD) was computed using PLINK.[73] From pairs with IBD coefficient ≥0.1875 we retained only one member. Individuals were restricted to those of EA by excluding outliers along the first two PCs from a principal component analysis seeded with the HapMap phase 3 release 2 populations as reference. The final dataset to estimate LD included 13,558 EA individuals and 16,969,363 SNPs.

The basis for statistical fine-mapping were the 256 1-Mb genome-wide significant loci identified in the EA meta-analysis, clipping at chromosome borders. Overlapping loci as well as pairs of loci whose respective index SNPs were correlated (r^2 >0.1 in the UKBB LD dataset described above) were merged. A single SNP was chosen to represent the MHC region, resulting in a final list of 212 regions prior to fine-mapping. Within each region, the GCTA COJO Slct algorithm[74] was used to identify independent variants employing a step-wise forward selection approach. We used the default collinearity cut-off of 0.9 (sensitivity analyses showing no major influence of alternative cutoff values; data not shown). We deemed an additional SNP as independently genome-wide significant if the SNPs' p-value conditional on all previously identified SNPs in the same region was $<5\times10^{-8}$.

Statistical fine-mapping and credible set generation in the EA meta-analysis

Statistical fine-mapping was carried out for each of the 212 regions. For each region containing multiple independent SNPs and for each independent SNP in such regions, approximate conditional analyses were carried out using the GCTA COJO-Cond algorithm to generate approximate conditional association statistics conditioned on the other independent SNPs in the region. Using the Wakefield's formula implemented in the R package 'gtx',[75] we derived approximate Bayes factors (ABF) from conditional estimates in regions with multiple independent SNPs and from the original estimates for regions with a single independent SNP. Given that 95% of the SNP effects on log(eGFR) fell within the -0.01 to 0.01 interval, the standard deviation prior was chosen as 0.0051 based on formula no. 8 in the original publication.[33] Sensitivity analyses showed that results were robust when higher values were used for the standard deviation prior (data not shown). For each variant within an evaluated region, the ABF obtained from the association betas and their standard errors of the marginal (single signal region) or conditional estimates (multi-signal regions) was used to calculate the posterior probability (PP) for the variant driving the association signal ("causal variant"). Ninety-nine percent credible sets, representing the set of SNPs that contain the causal

variant(s) with 99% probability, were computed by ranking variants by their PP and adding them to the set until the cumulative PP was >99% in each region.

Variant annotation

Functional annotation of variants mapping into credible sets was performed by querying the SNiPA database version 3.2 (March 2017),[76] based on the 1000Gp3v5 and Ensembl version 87 datasets. SNiPA was also used to derive the Combined Annotation Dependent Depletion (CADD) PHRED-like score,[77] based on CADD version 1.3. The Ensembl VEP tool[78] was used for SNP's primary effect prediction.

Co-localization analysis of associations with eGFR and gene expression (cis-eQTLs)

As the great majority of gene expression datasets is generated based on EA ancestry samples, co-localization analysis was based on the genetic associations with eGFR in the EA sample and with gene expression quantified from micro-dissected human glomerular and tubulo-interstitial kidney portions from 187 individuals participating in the NEPTUNE study,[47] as well as from the 44 tissues included in the GTEx Project version 6p release.[41] The eQTL and GWAS effect alleles were harmonized. For each locus, we identified tissue gene pairs with reported eQTL data within ±100 kb of each GWAS index variant. The region for each co-localization test was defined as the eQTL *cis* window defined in the underlying GTEx and NephQTL studies. We used the default parameters and prior definitions set in the 'coloc.fast' function from the R package 'gtx' (https://github.com/tobyjohnson/gtx), which is an adaption of Giambartolomei's colocalization method.[79] The package was also used to estimate the direction of effect over the credible sets as the ratio of the average PP weighted GWAS effects over the PP weighted eQTL effects.

Trans-eQTL analysis

We performed *trans*-eQTL annotation through LD mapping based on the 1000Gp3v5 European reference panel with an r^2 cut-off of >0.8. We limited annotation to index

SNPs with a fine-mapping posterior probability ≥1% in at least one fine-mapping-region. Due to expected small effect sizes, only genome-wide *trans*-eQTL studies of either peripheral blood mononuclear cells or whole blood with a sample size of ≥1000 individuals were considered, resulting in five non-overlapping studies[80-84] (**Supplementary Table 15**). For the study by Kirsten *et al*,[84] we had access to an update with larger sample size combining two non-overlapping studies (LIFE-Heart[85] and LIFE-Adult[86]) resulting in a total sample size of 6645. To improve stringency of results, we focused the analysis on inter-chromosomal *trans*-eQTLs with $P<5\times10^{-8}$ reported by ≥2 studies.

Co-localization analyses with urinary uromodulin concentrations

Association between concentrations of the urinary uromodulin-to-creatinine ratio with genetic variants at the *UMOD-PDILT* locus were evaluated in the German Chronic Kidney Disease (GCKD) study.[87] Uromodulin concentrations were measured from frozen stored urine using an established ELISA assay with excellent performance as described previously.[37] Concentrations were indexed to creatinine to account for urine dilution. Genetic associations were computed using the same software and settings as for the association with eGFR (**Supplementary Table 2**). Co-localization analyses were carried out using identical software and settings as described above for the association with gene expression.

Acknowledgements

We thank Daniele Di Domizio (Eurac Research) and Jochen Knaus (University of Freiburg) for IT assistance, and Toby Johnson (GSK) for sharing his code and discussion on credible set fine-mapping and co-localization analysis. This research has been conducted using the UK Biobank Resource under Application Number 20272. Study-specific acknowledgements and funding sources are listed in the **Supplementary Material** (page 12).

Disclaimer

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the US Department of Health and Human Services.

Author contributions

Manuscript writing group: Matthias Wuttke, Yong Li, Man Li, Karsten Sieber, Mary Feitosa, Mathias Gorski, Adrienne Tin, Lihua Wang, Holger Kirsten, Tarunveer Ahluwalia, Kevin Ho, Iris Heid, Markus Scholz, Alexander Teumer, Anna Köttgen, Cristian Pattaro

Design of this study: Carsten A. Böger, Christian Fuchsberger, Mathias Gorski, Anna Köttgen, Andrew P. Morris, Cristian Pattaro, Alexander Teumer, Adrienne Tin, Matthias Wuttke

Management of an individual contributing study: Tarunveer S. Ahluwalia, Emanuele di Angelantonio, Shreeram Akilesh, Stephan J.L. Bakker, Ginevra Biino, Murielle Bochud, Michael Boehnke, Eric Boerwinkle, Martin H. de Borst, Hermann Brenner, Adam S. Butterworth, Carsten A. Böger, Archie Campbell, Robert J. Carroll, John C. Chambers, Daniel I. Chasman, Ching-Yu Cheng, Kaare Christensen, Renata Cifkova, Marina Ciullo, Josef Coresh, Daniele Cusi, Rob M. van Dam, John Danesh, Olivier Devuyst, Cornelia M. van Duijn, Kai-Uwe Eckardt, Georg Ehret, Paul Elliott, Michele K. Evans, Janine F. Felix, Oscar H. Franco, Barry I. Freedman, Yechiel Friedlander, Ron T. Gansevoort, He Gao, Paolo Gasparini, J. Michael Gaziano, Vilmantas Giedraitis, Christian Gieger, Franco Giulianini, Alessandro De Grandi, Vilmundur Gudnason, Tamara B. Harris, Pim van der Harst, Catharina A. Hartman, Caroline Hayward, Chew-Kiat Heng, Andrew A. Hicks, Kevin Ho, Adriana Hung, M. Arfan Ikram, Olafur S. Indridason, Erik Ingelsson, Vincent W.V. Jaddoe, Jost B. Jonas, Bettina Jung, Candace M. Kammerer, Chiea Chuen Khor, Wieland Kiess, Marcus E. Kleber, Wolfgang Koenig, Jaspal S. Kooner, Holly Kramer, Florian Kronenberg, Bernhard K. Krämer, Michiaki Kubo, Johanna Kuusisto, Mika Kähönen, Antje Körner, Anna Köttgen, Terho Lehtimäki, Yong Li, Su-Chi Lim, Markus Loeffler, Ruth J.F. Loos, Susanne Lucae, Mary Ann Lukas, Patrik K.E. Magnusson, Nicholas G. Martin, Deborah Mascalzoni, Koichi Matsuda, Olle Melander, Andres Metspalu, Evgenia K. Mikaelsdottir, Yuri Milaneschi, Karen L. Mohlke, Grant W. Montgomery, Andrew P. Morris, Renée de Mutsert, Winfried März, Girish N. Nadkarni, Jeffrey O'Connell, Michelle L. O'Donoghue, Albertine J. Oldehinkel, Marju Orho-Melander, Willem H. Ouwehand, Afshin Parsa, Cristian Pattaro, Sarah A. Pendergrass, Brenda W.J.H. Penninx, Thomas Perls, Markus Perola, Mario Pirastu, Ozren Polasek, Belen Ponte, Peter P. Pramstaller, Michael A. Province, Bruce M. Psaty, Ton J. Rabelink, Olli T. Raitakari, Dermot F. Reilly, Rainer Rettig, Myriam Rheinberger, Paul M. Ridker, David J. Roberts, Peter Rossing, Igor Rudan, Charumathi Sabanayagam, Veikko Salomaa, Kai-Uwe Saum, Helena Schmidt, Reinhold Schmidt, Markus Scholz, Ben Schöttker, Xueling Sim, Harold Snieder, Nicole Soranzo, Cassandra N. Spracklen, Kari Stefansson, Konstantin Strauch, Michael Stumvoll, Gardar Sveinbjornsson, Per O. Svensson, E-Shyong Tai, Bamidele O. Tayo, Yih-Chung Tham, Joachim Thiery, Adrienne Tin, Daniela Toniolo, Johanne Tremblay, Ioanna Tzoulaki, Anke Tönjes, Peter Vollenweider, Aiko P.J. de Vries, Uwe Völker, Gerard Waeber, Lars Wallentin, Ya Xing Wang, Dawn M. Waterworth, Wen Bin Wei, Harvey White, John B. Whitfield, Sarah H. Wild, James G. Wilson, Charlene Wong, Tien Yin Wong, Matthias Wuttke, Liang Xu, Qiong Yang, Masayuki Yasuda, Weihua Zhang, Alan B. Zonderman

Critical review of manuscript: Tarunveer S. Ahluwalia, Shreeram Akilesh, Peter Almgren, Emanuele di Angelantonio, Stephan J.L. Bakker, Nisha Bansal, Mary L. Biggs, Ginevra Biino, Martin H. de Borst, Erwin P. Bottinger, Thibaud S. Boutin, Hermann Brenner, Adam S. Butterworth, Carsten A. Böger, Harry Campbell, Daniel I. Chasman, Xu Chen, Yurong Cheng, Audrey Y. Chu, Marina Ciullo, Josef Coresh, Rob M. van Dam, Graciela Delgado, Olivier Devuyst, Jasmin Divers, Rajkumar Dorajoo, Kai-Uwe Eckardt, Digna R. Velez Edward, Todd L. Edwards, Paul Elliott, Karlhans Endlich, Michele K. Evans, Mary F. Feitosa, Janine F. Felix, Oscar H. Franco, Andre Franke, Barry I. Freedman, Yechiel Friedlander, Christian Fuchsberger, He Gao, Sahar Ghasemi, Christian Gieger, Ayush Giri, Scott D. Gordon, Mathias Gorski, Daniel F. Gudbjartsson, Pavel Hamet, Tamara B. Harris, Pim van der Harst, Catharina A. Hartman, Caroline Hayward, Iris M. Heid, Jacklyn N. Hellwege, Chew-Kiat Heng, Kevin Ho, Anselm Hoppmann, Wei Huang, Nina Hutri-Kähönen, Shih-Jen Hwang, Olafur S. Indridason, Erik Ingelsson, Vincent W.V. Jaddoe, Johanna Jakobsdottir, Jost B. Jonas, Peter K. Joshi, Bettina Jung, Mika Kastarinen, Shona M. Kerr, Marcus E. Kleber, Wolfgang Koenig, Aldi T. Kraja, Holly Kramer, Florian Kronenberg, Bernhard K. Krämer, Mikko Kuokkanen, Mika Kähönen, Antje Körner, Anna Köttgen, Brigitte Kühnel, Markku Laakso, Leslie A. Lange, Carl D. Langefeld, Jeannette Jen-Mai Lee, Terho Lehtimäki, Man Li, Yong Li, Wolfgang Lieb, Lars Lind, Cecilia M. Lindgren, Markus Loeffler, Ruth J.F. Loos, Leo-Pekka Lyytikäinen, Patrik K.E. Magnusson, Anubha Mahajan, Jonathan Marten, Nicholas G. Martin, Deborah Mascalzoni, Christa Meisinger, Thomas Meitinger, Olle Melander, Evgenia K. Mikaelsdottir, Kozeta Miliku, Karen L. Mohlke, Grant W. Montgomery, Dennis O. Mook-Kanamori, Renée de Mutsert, Winfried März, Girish N. Nadkarni, Mike A. Nalls, Matthias Nauck, Kjell Nikus, Boting Ning, Ilja M. Nolte, Raymond Noordam, Teresa Nutile, Michelle L. O'Donoghue, Albertine J. Oldehinkel, Marju Orho-Melander, Nicholette D. Palmer, Runolfur Palsson, Afshin Parsa, Cristian Pattaro, Sarah A. Pendergrass, Brenda W.J.H. Penninx, Markus Perola, Ozren Polasek, Michael H. Preuss, Bram P. Prins, Bruce M. Psaty, Ton J. Rabelink, Laura M. Raffield, Olli T. Raitakari, Rainer Rettig, Myriam Rheinberger, Kenneth M. Rice, Paul M. Ridker, Fernando Rivadeneira, David J. Roberts, Peter Rossing, Igor Rudan, Daniela Ruggiero, Charumathi Sabanayagam, Veikko Salomaa, Kai-Uwe Saum, Markus Scholz, Christina-Alexandra Schulz, Nicole Schupf, Ben Schöttker, Sanaz Sedaghat, Karsten B. Sieber, Xueling Sim, Albert V. Smith, Harold Snieder, Cassandra N. Spracklen, Konstantin Strauch, Gardar Sveinbjornsson, Per O. Svensson, Salman M. Tajuddin, Nicholas Y. Q. Tan, Bamidele O. Tayo, Alexander Teumer, Hauke Thomsen, Adrienne Tin, Johanne Tremblay, Ioanna Tzoulaki, Anke Tönjes, André G. Uitterlinden, Niek Verweij, Veronique

Vitart, Suzanne Vogelezang, Aiko P.J. de Vries, Uwe Völker, Melanie Waldenberger, Lars Wallentin, Dawn M. Waterworth, Harvey White, John B. Whitfield, Sarah H. Wild, James G. Wilson, Matthias Wuttke, Qiong Yang, Zhi Yu, Alan B. Zonderman

Statistical Methods and Analysis: Tarunveer S. Ahluwalia, Masato Akiyama, Peter Almgren, Mary L. Biggs, Ginevra Biino, Mathilde Boissel, Thibaud S. Boutin, Marco Brumat, Carsten A. Böger, Mickaël Canouil, Robert J. Carroll, Jin-Fang Chai, Daniel I. Chasman, Miao-Li Chee, Xu Chen, Yurong Cheng, Audrey Y. Chu, Massimiliano Cocca, Maria Pina Concas, James P. Cook, Tanguy Corre, Abbas Dehghan, Graciela Delgado, Ayse Demirkan, Jasmin Divers, Rajkumar Dorajoo, Digna R. Velez Edward, Todd L. Edwards, Mary F. Feitosa, Janine F. Felix, Barry I. Freedman, Sandra Freitag-Wolf, Christian Fuchsberger, Sahar Ghasemi, Ayush Giri, Mathias Gorski, Daniel F. Gudbjartsson, Martin Gögele, Toomas Haller, Pavel Hamet, Pim van der Harst, Iris M. Heid, Jacklyn N. Hellwege, Edith Hofer, Anselm Hoppmann, Katrin Horn, Shih-Jen Hwang, Johanna Jakobsdottir, Peter K. Joshi, Navya Shilpa Josyula, Bettina Jung, Yoichiro Kamatani, Masahiro Kanai, Chiea-Chuen Khor, Holger Kirsten, Marcus E. Kleber, Alena Krajcoviechova, Holly Kramer, Mikko Kuokkanen, Anna Köttgen, Brigitte Kühnel, Leslie A. Lange, Carl D. Langefeld, Man Li, Yong Li, Jianjun Liu, Jun Liu, Leo-Pekka Lyytikäinen, Anubha Mahajan, Jonathan Marten, Jade Martins, Kozeta Miliku, Pashupati P. Mishra, Nina Mononen, Andrew P. Morris, Peter J. van der Most, Winfried März, Mike A. Nalls, Matthias Nauck, Boting Ning, Damia Noce, Ilja M. Nolte, Raymond Noordam, Teresa Nutile, Yukinori Okada, Cristian Pattaro, Sarah A. Pendergrass, Nicola Pirastu, Michael H. Preuss, Bram P. Prins, Laura M. Raffield, Myriam Rheinberger, Kenneth M. Rice, Fernando Rivadeneira, Federica Rizzi, Rico Rueedi, Kathleen A. Ryan, Yasaman Saba, Erika Salvi, Markus Scholz, Christina-Alexandra Schulz, Sanaz Sedaghat, Yuan Shi, Karsten B. Sieber, Xueling Sim, Albert V. Smith, Cassandra N. Spracklen, Heather M. Stringham, Gardar Sveinbjornsson, Silke Szymczak, Salman M. Tajuddin, Bamidele O. Tayo, Alexander Teumer, Chris H.L. Thio, Hauke Thomsen, Gudmar Thorleifsson, Johanne Tremblay, Niek Verweij, Veronique Vitart, Suzanne Vogelezang, Chaolong Wang, Lihua Wang, James F. Wilson, Mary K. Wojczynski, Matthias Wuttke, Yizhe Xu, Qiong Yang, Laura M. Yerges-Armstrong, Weihua Zhang

Subject Recruitment: Saima Afaq, Erwin P. Bottinger, Hermann Brenner, Carsten A. Böger, Archie Campbell, Harry Campbell, John C. Chambers, Miao-Ling Chee, Kaare Christensen, Renata Cifkova, Marina Ciullo, Daniele Cusi, Katalin Dittrich, Michele K. Evans, Valencia Hui Xian Foo, Barry I. Freedman, Ron T. Gansevoort, Vilmundur Gudnason, Catharina A. Hartman, Wei Huang, Nina Hutri-Kähönen, Olafur S. Indridason, Marcus Ising, Vincent W.V. Jaddoe, Jost B. Jonas, Bettina Jung, Candace M. Kammerer, Mika Kastarinen, Jaspal S. Kooner, Alena Krajcoviechova, Florian Kronenberg, Michiaki Kubo, Mika Kähönen, Anna Köttgen, Markku Laakso, Jeannette Jen-Mai Lee, Terho Lehtimäki, Wolfgang Lieb, Lars Lind, Nicholas G. Martin, Koichi Matsuda, Christa Meisinger, Andres Metspalu, Renée de Mutsert, Winfried März, Kjell Nikus, Michelle L. O'Donoghue, Isleifur Olafsson, Albertine J. Oldehinkel, Sandosh Padmanabhan, Cristian Pattaro, Sarah A. Pendergrass, Brenda W.J.H. Penninx, Markus Perola, Ozren Polasek, Belen Ponte, David J. Porteous, Tanja Poulain, Michael A. Province, Ton J. Rabelink, Olli T. Raitakari, Myriam Rheinberger, Paul M. Ridker, Peter Rossing, Igor Rudan, Daniela Ruggiero, Veikko Salomaa, Reinhold Schmidt, Blair H. Smith, Per O. Svensson, Nicholas Y. Q. Tan, Andrej Teren, Yih-Chung Tham, Johanne Tremblay, Ioanna Tzoulaki, Anke Tönjes, Simona Vaccargiu, Suzanne Vogelezang, Peter Vollenweider, Aiko P.J. de Vries, Gerard Waeber, Lars Wallentin, Harvey White, John B. Whitfield, Sarah H. Wild, James G. Wilson, Alan B. Zonderman, Johan Ärnlöv

Bioinformatics: Tarunveer S. Ahluwalia, Shreeram Akilesh, Peter Almgren, Daniela Baptista, Sven Bergmann, Adam S. Butterworth, Carsten A. Böger, Eric Campana, Robert J. Carroll, Xu Chen, Audrey Y. Chu, Massimiliano Cocca, Maria Pina Concas, Tanguy Corre, E. Warwick Daw, Frauke Degenhardt, Abbas Dehghan, Jasmin Divers, Rajkumar Dorajoo, Georg Ehret, Andre Franke, He Gao, Sahar Ghasemi, Ayush Giri, Scott D. Gordon, Mathias Gorski, Pavel Hamet, Iris M. Heid, Edith Hofer, Anselm Hoppmann, Katrin Horn, Johanna Jakobsdottir, Navya Shilpa Josyula, Chiea-Chuen Khor, Holger Kirsten, Marcus E. Kleber, Alena Krajcoviechova, Anna Köttgen, Carl D. Langefeld, Benjamin Lehne, Man Li, Yong Li, Jianjun Liu, Leo-Pekka Lyytikäinen, Jonathan Marten, Jade Martins, Yuri Milaneschi, Pashupati P. Mishra, Karen L. Mohlke, Dennis O. Mook-Kanamori, Peter J. van der Most, Reedik Mägi, Winfried März, Raymond Noordam, Teresa Nutile, Sarah A. Pendergrass, Nicola Pirastu, Giorgio Pistis, Anna I. Podgornaia, Michael H. Preuss, Bram P. Prins, Federica Rizzi, Rico Rueedi, Yasaman Saba, Erika Salvi, Markus Scholz, Christina-Alexandra Schulz, Sanaz Sedaghat, Christian M. Shaffer, Karsten B. Sieber, Albert V. Smith, Cassandra N. Spracklen, Silke Szymczak, Hauke Thomsen, Johanne Tremblay, Chaolong Wang, James F. Wilson, Matthias Wuttke, Yizhe Xu, Laura M. Yerges-Armstrong, Zhi Yu, Weihua Zhang

Interpretation of Results: Tarunveer S. Ahluwalia, Emanuele di Angelantonio, Carsten A. Böger, Ching-Yu Cheng, Katalin Dittrich, Jasmin Divers, Rajkumar Dorajoo, Karlhans Endlich, Mary F. Feitosa, Janine F. Felix, Barry I. Freedman, Sahar Ghasemi, Christian Gieger, Ayush Giri, Mathias Gorski, Pavel Hamet, Pim van der Harst, Hauke Thomsen, Iris M. Heid, Kevin Ho, Katrin Horn, Wei Huang, Shih-Jen Hwang, Bettina Jung, Holger Kirsten, Wolfgang Koenig, Alena Krajcoviechova, Anna Köttgen, Markku Laakso, Carl D. Langefeld, Man Li, Yong Li, Patrik K.E. Magnusson, Jonathan Marten, Kozeta Miliku, Karen L. Mohlke, Andrew P. Morris, Nicholette D. Palmer, Cristian Pattaro, Sarah A. Pendergrass, Bram P. Prins, Dermot F. Reilly, Myriam Rheinberger, Paul M. Ridker, Markus Scholz, Sanaz Sedaghat, Karsten B. Sieber, Cassandra N. Spracklen, Per O. Svensson, Bamidele O. Tayo,

Alexander Teumer, Adrienne Tin, Johanne Tremblay, Ioanna Tzoulaki, André G. Uitterlinden, Niek Verweij, Veronique Vitart, Suzanne Vogelezang, Lars Wallentin, Harvey White, Matthias Wuttke, Yizhe Xu, Masayuki Yasuda, Laura M. Yerges-Armstrong

Genotyping: Najaf Amin, Daniela Baptista, Ralph Burkhardt, Adam S. Butterworth, Carsten A. Böger, Archie Campbell, Harry Campbell, Daniel I. Chasman, Ching-Yu Cheng, E. Warwick Daw, Ayse Demirkan, Rajkumar Dorajoo, Cornelia M. van Duijn, Georg Ehret, Michele K. Evans, Mary F. Feitosa, Andre Franke, Yechiel Friedlander, Christian Fuchsberger, Ron T. Gansevoort, He Gao, Scott D. Gordon, Pavel Hamet, Pim van der Harst, Hauke Thomsen, Caroline Hayward, Chew-Kiat Heng, Wei Huang, Erik Ingelsson, Chiea Chuen Khor, Marcus E. Kleber, Wolfgang Koenig, Jaspal S. Kooner, Peter Kovacs, Aldi T. Kraja, Alena Krajcoviechova, Florian Kronenberg, Michiaki Kubo, Mika Kähönen, Antje Körner, Leslie A. Lange, Terho Lehtimäki, Leo-Pekka Lyytikäinen, Patrik K.E. Magnusson, Thomas Meitinger, Olle Melander, Yuri Milaneschi, Karen L. Mohlke, Nina Mononen, Grant W. Montgomery, Dennis O. Mook-Kanamori, Andrew P. Morris, Josyf C. Mychaleckyj, Winfried März, Mike A. Nalls, Marju Orho-Melander, Sandosh Padmanabhan, Nicholette D. Palmer, Brenda W.J.H. Penninx, Markus Perola, David J. Porteous, Michael H. Preuss, Olli T. Raitakari, Dermot F. Reilly, Fernando Rivadeneira, Federica Rizzi, Jerome I. Rotter, Daniela Ruggiero, Veikko Salomaa, Erika Salvi, Blair H. Smith, Cassandra N. Spracklen, Salman M. Tajuddin, Kent Taylor, Alexander Teumer, Daniela Toniolo, Johanne Tremblay, André G. Uitterlinden, Simona Vaccargiu, Uwe Völker, Melanie Waldenberger, Chaolong Wang, Lihua Wang, Ya Xing Wang, James G. Wilson, Mary K. Wojczynski, Alan B. Zonderman, Johan Ärnlöv

Competing interests

Wolfgang Koenig reports modest consultation fees for advisory board meetings from Amgen, DalCor, Kowa, Novartis, Pfizer and Sanofi, and modest personal fees for lectures from Amgen, AstraZeneca, Novartis, Pfizer and Sanofi, all outside the submitted work. Winfried März is employed with Synlab Services GmbH and holds shares of Synlab Holding Deutschland GmbH. Dennis O. Mook-Kanamori is a part time research consultant at Metabolon Inc. Mike A. Nalls is supported by a consulting contract between Data Tecnica International LLC and the National Institute on Aging (NIA), National Institutes of Health (NIH), Bethesda, MD, USA and consults for Illumina Inc., the Michael J. Fox Foundation, and the University of California Healthcare. Oscar H. Franco works in ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.); Metagenics Inc.; and AXA. Karsten B. Sieber, Laura Yerges-Armstrong, Dawn M. Waterworth, and Mary Ann Lukas, are full-time employees of GlaxoSmithKline. Michelle L. O'Donoghue received grant support from GlaxoSmithKline, MSD, Eisai, AstraZeneca, MedCo and Janssen. Harvey White: Grants and non-financial support from GlaxoSmithKline, during the conduct of the study; grants from Sanofi-Aventis, Eli Lilly, National Institute of Health, George Institute, Omthera Pharmaceuticals, Pfizer New Zealand, Intarcia Therapeutics Inc., Elsai Inc., Dalcor Pharma UK Inc,; Honoraria and non-financial support from AstraZeneca and; Advisory Boards for Sirtex and Acetilion and personal fees from CSL Behring LLC and Luitpold Pharmaceuticals Inc. outside the submitted work. Lars Wallentin: Institutonal grants from GlaxoSmithKline, AstraZeneca, BMS, Boehringer-Ingelheim, Pfizer. MSD, Roche Diagnostics. Dermot F. Reilly and Anna I. Podgornaia are employees of Merck Sharp Dohme Corp., Whitehouse Station, NJ, USA. Markus Scholz: Consultancy of and grant support from Merck Serono not related to this project. Bruce M. Psaty serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. John Danesh is member of the Novartis Cardiovascular and Metabolic Advisory Board, received grant support from Novartis. Adam S. Butterworth received grants from MSD, Pfizer, Novartis, Biogen and Bioverativ and personal fees from Novartis. Veikko Salomaa has participated in a conference trip sponsored by Novo Nordisk and received

a honorarium from the same source for participating in an advisory board meeting. All other authors declare no conflicts of interest.

Figure Legends

Figure 1 – Trans-ethnic GWAS meta-analysis identifies 308 loci associated with eGFR

Circos plot: Red band: $-\log_{10}(P)$ for association with eGFR, by chromosomal position. Blue line indicates genome-wide significance ($P=5\times10^{-8}$). Black gene labels indicate novel loci, blue labels known loci. Non-replicating loci are colored in gray (novel) or light blue (known). Green band: Measures of heterogeneity related to the index SNPs associated with eGFR. Dot sizes are proportional to I^2 or ancestry-related heterogeneity (p-anc-het). Blue band: $-\log_{10}(P)$ for association with CKD, by chromosomal position. Red line indicates genome-wide significance ($P=5\times10^{-8}$). Radial lines mark regions with p-anc-het $<1.6\times10^{-4}$ = 0.05/308 or I^2 >25%. Inset: Effects of all 308 index SNPs on log(eGFR) by their minor allele frequency, color-coded by the associated odds ratio (OR) of CKD (red scale for OR≤1, blue scale for OR>1). Triangles highlight SNPs that were significantly ($P<1.6\times10^{-4}$ =0.05/308) associated with CKD.

Figure 2 – Generalizability with respect to other populations and other kidney function markers

Panel A: Measures of heterogeneity for 308 eGFR-associated index SNPs. Comparison of each variant's heterogeneity quantified as I^2 from the trans-ethnic meta-analysis (Y-axis) vs. ancestry-related heterogeneity from meta-regression ($-\log_{10}$(p-anc-het), X-axis). Histograms summarize the distribution of the heterogeneity measures on both axes. SNPs with ancestry-related heterogeneity (p-anc-het$<1.6\times10^{-4}$ = 0.05/308) are marked in blue and labeled; SNPs with I^2>50% are labeled. **Panel B: Comparison of genetic effect sizes between CKDGen Consortium data (X-axis) and MVP data (Y-axis).** Blue font indicates $P<1.6\times10^{-4}$ (0.05/308) in the MVP. Error bars indicate 95% CIs. Dashed line: line of best fit. **Panel C: Comparison of the magnitude of the effects on eGFR (X-axis) vs. BUN (Y-axis) for the 308 eGFR-associated index SNPs.** SNPs are marked in blue when $P<1.6\times10^{-4}$ (0.05/308) in the BUN analysis. Error

bars indicate 95% CIs. Dashed line: line of best fit. **Panel D**: **Association between a genetic risk score (GRS) for low eGFR and clinical phenotypes abstracted from ICD-10 codes independent individuals from the UK Biobank**. Asthma was included as a negative control. Displayed are odds ratios and their 95% CIs per standard deviation increase in GRS (Methods).

Figure 3 – Human orthologs of genes with renal phenotypes in genetically manipulated mice are enriched for association signals with eGFR

Signals in candidate genes identified based on the murine phenotypes abnormal GFR (**Panel A**), abnormal kidney physiology (**Panel B**), and abnormal kidney morphology (**Panel C**). *Y*-axis: $-\log_{10}(P)$ for association with eGFR in the trans-ethnic meta-analysis for the variant with the lowest p-value in each candidate gene. Dashed line indicates genome-wide significance ($P=5\times10^{-8}$), solid gray line indicates the experiment-wide significance threshold for each nested candidate gene analysis (included in lower right corner in each panel). Orange color indicates genome-wide significance, red color experiment-wide but not genome-wide significance, and blue color indicates genes with no significantly associated SNPs. Genes are labeled when reaching experiment- but not genome-wide significance; black font for genes not mapping into loci reported in the main analysis, gray font otherwise. Enrichment p-value reported for observed number of genes with association signals below the experiment-wide threshold against the expected number based on the complementary cumulative binomial distribution (Methods).

Figure 4 – Credible set size (X-axis) against variant posterior probability (Y-axis) of 3655 variants in 252 99% credible sets by annotation

Panel A: Exonic variants. Variants are marked by triangles, with size proportional to their CADD score. Red triangles and variant labeling indicate missense variants mapping into small (≤5 SNPs) credible sets or with high individual posterior probability of driving the association signal (>0.5). **Panel B: Regulatory potential.** Symbol colors identify variants with regulatory potential as derived from DNAse hypersensitivity analysis in target tissues (Methods). Variant annotation was restricted to variants with variant posterior probability >1%; SNPs with posterior probability ≥90% contained in credible sets with ≤10 variants were labeled.

Figure 5 – Co-localization of eGFR-association signals with gene expression in kidney tissues

All eGFR loci were tested for co-localization with all eQTLs where the eQTL cis-window overlapped (±100 kb) the sentinel genetic variants. Genes with ≥1 positive co-localization (posterior probability of one common causal variant, H4, ≥0.80) in a kidney tissue are illustrated with the respective sentinel variants (Y-axis). Co-localizations across all tissues (X-axis) are illustrated as dots, where the size of the dots indicates the posterior probability of the co-localization. Negative co-localizations (posterior probability of H4 <0.80) are marked in grey, while the positive co-localizations are color-coded based on the predicted change in expression relative to the allele associated with lower eGFR.

Figure 6 – Co-localization of independent eGFR-association signals at the *UMOD-PDILT* locus with urinary uromodulin concentrations supports *UMOD* as the effector gene.

Association plots: association $-\log_{10}$(p-value) (Y axis) vs. chromosomal position (X axis). Approximate conditional analyses among EA individuals support the presence of two independent eGFR-associated signals (**Panel A**). The association signal with urinary uromodulin/creatinine levels looks similar (**Panel B**). Co-localization of association with eGFR (upper sub-panel) and urinary uromodulin/creatinine levels (lower sub-panel) for the independent regions centered on *UMOD* (**Panel C**) and *PDILT* (**Panel D**) support a shared underlying variant in both regions with high posterior probability.

Table 1 – Genes implicated as causal via identification of missense variants with high probability of driving the eGFR association signal. Genes are included if they contain a missense variant with posterior probability of association of >50% or mapping into a small credible set (≤5 variants).

Gene	SNP	Credible set size	SNP PP[1]	functional consequence	CADD score[2]	DHS[3], tissue	**Brief summary of the gene's function and relevant literature** (OMIM entries are indicated as #number)
CACNA1S	rs3850625	1	1.00	p.Arg1539Cys (NP_000060.2)	34.0	-	Encodes a subunit of the slowly inactivating L-type voltage-dependent calcium channel in skeletal muscle. Reports of altered expression in kidney cancer (PMID 28781648) and after indoxyl sulfate treatment (PMID: 27550174). Rare variants can cause autosomal dominant hypokalemic periodic paralysis, type 1 (#170400) or malignant hyperthermia susceptibility (#601887). Common variation at this locus has been reported as associated with eGFR in previous GWAS (PMID: 24029420, PMID: 26831199).
EDEM3	rs78444298	1	1.00	p.Pro746Ser (NP_079467.3)	24.6	-	The gene product accelerates the glycoprotein ER-associated degradation by proteasomes by catalyzing mannose trimming from Man8GlcNAc2 to Man7GlcNAc2 in the N-glycans. This variant has been identified by a previous exome chip association study with eGFR (PMID: 27920155).
KLHDC7A	rs11261022	7	0.71	p.Arg160Ser (NP_689588.2)	1.1	Roadmap, ENCODE kidney	Kelch Domain Containing 7A is a protein coding gene and a paralog of *KBTBD11*. No specific entry in relation to kidney disease in PubMed.
RPL3L	rs113956264	1	1.00	p.Val262Met (NP_005052.1)	27.2	-	The gene product shares sequence similarity with ribosomal protein L3. It has a tissue-specific expression pattern, with highest levels in skeletal muscle and heart.
SLC25A45	rs34400381	1	1.00	p.Arg285Cys (NP_001070709.2)	26.0	ENCODE kidney	Belongs to the SLC25 family of mitochondrial carrier proteins and is an orphan transporter. This variant has already been identified in a GWAS of symmetric dimethylarginine levels (PMID: 24159190) and in a whole-genome sequence (WGS) analysis of serum creatinine (PMID: 25082825). *SLC25A45* may play a role in biosynthesis of arginine, which is involved in the synthesis of creatine.
SLC47A1	rs111653425	1	1.00	p.Ala465Val (NP_060712.2)	24.6	-	Encodes the multidrug and toxin extrusion protein (MATE1), a transport protein responsible for the secretion of cationic drugs and creatinine across brush border membranes. This variant has already been identified in a WGS analysis of serum creatinine from Iceland (PMID: 25082825). Rare and common variants in the locus have been identified in exome chip (PMID: 27920155) and in GWAS (PMID: 20383146) studies of eGFR, respectively. MATE1 knockout (KO) mice show higher levels of serum creatinine and BUN (PMID: 19332510), arguing against a sole effect on creatinine transport and supporting an effect on kidney function.
PPM1J	rs34611728	5	0.02	p.Leu213Phe (NP_005158.5)	13.1	ENCODE kidney	This gene encodes the serine/threonine protein phosphatase. The variant has been reported in association with eGFR in an exome chip association study (PMID: 27920155).
CERS2	rs267738	5	0.46	p.Glu115Ala (NP_071358.1)	32.0/ 28.2	-	Encodes Ceramide Synthase 2, which may be involved in sphingolipid synthesis. Changes in ceramides were reported as essential in renal Madin-Darby Canine Kidney (MDCK) cell differentiation (PMID: 28515139). *CERS2* KO mice show strongly reduced ceramide levels in the kidney and develop renal parenchyma abnormalities (PMID: 19801672). This variant has been reported as associated with the rate of albuminuria increase in patients with diabetes (PMID: 25238615).
C9	rs700233	5	0.32	p.Arg5Trp (NP_001728.1)	6.6	-	Encodes a constituent of the membrane attack complex that plays a key role in the innate and adaptive immune response. Rare mutations can cause C9 deficiency (#613825). *C9* is mentioned in several kidney disease case reports, including patients with congenital factor 9 deficiency showing IgA nephropathy (PMID: 1453611).
SLC22A2	rs316019	4	0.04	p.Ser270Ala (NP_003049.2)	12.7	-	Encodes the polyspecific organic cation transporter (OCT2) that is primarily expressed in the kidney, where it mediates tubular uptake of organic compounds including creatinine from the circulation. Many publications relate *SLC22A2* to kidney function. rs316019 is a known pharmacogenomics variant associated with response to metformin and other drugs such as cisplatin. Carriers of the risk allele have a higher risk of cisplatin-induced nephrotoxicity (PMID: 19625999), indicating that this transporter is essential in excreting toxins. The locus has been reported in previous GWAS of eGFR (PMID: 20383146).

[1]PP: posterior probability. [2]CADD score: Combined Annotation Dependent Depletion (CADD) PHRED-like score (Methods); [3]DHS: DNAse Hypersensitivity Site

References

1. Eckardt, K.U. *et al.* Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet* **382**, 158-69 (2013).
2. Jha, V. *et al.* Chronic kidney disease: global dimension and perspectives. *Lancet* **382**, 260-72 (2013).
3. Ene-Iordache, B. *et al.* Chronic kidney disease and cardiovascular risk in six regions of the world (ISN-KDDC): a cross-sectional study. *Lancet Glob Health* **4**, e307-19 (2016).
4. Go, A.S., Chertow, G.M., Fan, D., McCulloch, C.E. & Hsu, C.Y. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* **351**, 1296-305 (2004).
5. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1151-1210 (2017).
6. Inrig, J.K. *et al.* The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials.gov. *Am J Kidney Dis* **63**, 771-80 (2014).
7. Levin, A. *et al.* Global kidney health 2017 and beyond: a roadmap for closing gaps in care, research, and policy. *Lancet* **390**, 1888-1917 (2017).
8. Wuttke, M. & Kottgen, A. Insights into kidney diseases from genome-wide association studies. *Nat Rev Nephrol* **12**, 549-62 (2016).
9. Gorski, M. *et al.* 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Sci Rep* **7**, 45040 (2017).
10. Pattaro, C. *et al.* Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun* **7**, 10023 (2016).
11. Chasman, D.I. *et al.* Integration of genome-wide association studies with biological knowledge identifies six novel genes related to kidney function. *Hum Mol Genet* **21**, 5329-43 (2012).
12. Pattaro, C. *et al.* Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet* **8**, e1002584 (2012).
13. Kottgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42**, 376-84 (2010).
14. Chambers, J.C. *et al.* Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet* **42**, 373-5 (2010).
15. Kottgen, A. *et al.* Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet* **41**, 712-7 (2009).
16. Okada, Y. *et al.* Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet* **44**, 904-9 (2012).
17. Hishida, A. *et al.* Genome-Wide Association Study of Renal Function Traits: Results from the Japan Multi-Institutional Collaborative Cohort Study. *Am J Nephrol* **47**, 304-316 (2018).
18. Lee, J. *et al.* Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLoS One* **13**, e0194044 (2018).
19. Mahajan, A. *et al.* Trans-ethnic Fine Mapping Highlights Kidney-Function Genes Linked to Salt Sensitivity. *Am J Hum Genet* **99**, 636-646 (2016).
20. Devuyst, O. & Pattaro, C. The UMOD Locus: Insights into the Pathogenesis and Prognosis of Kidney Disease. *J Am Soc Nephrol* **29**, 713-726 (2018).
21. Yeo, N.C. *et al.* Shroom3 contributes to the maintenance of the glomerular filtration barrier integrity. *Genome Res* **25**, 57-65 (2015).
22. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390-400 (2018).
23. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).

24. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
25. Gudbjartsson, D.F. *et al.* Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases. *PLoS Genet* **6**, e1001039 (2010).
26. Magi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* **26**, 3639-3650 (2017).
27. Gaziano, J.M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-23 (2016).
28. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
29. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
30. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
31. Jing, J. *et al.* Combination of mouse models and genomewide association studies highlights novel genes associated with human kidney function. *Kidney Int* **90**, 764-73 (2016).
32. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539-551 (2017).
33. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
34. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **24**, 2125-37 (2015).
35. Tsuda, M. *et al.* Targeted disruption of the multidrug and toxin extrusion 1 (mate1) gene in mice reduces renal secretion of metformin. *Mol Pharmacol* **75**, 1280-6 (2009).
36. Karczewski, K.J. & Snyder, M.P. Integrative omics for health and disease. *Nat Rev Genet* **19**, 299-310 (2018).
37. Olden, M. *et al.* Common variants in UMOD associate with urinary uromodulin levels: a meta-analysis. *J Am Soc Nephrol* **25**, 1869-82 (2014).
38. Moreau, M.E. *et al.* The kallikrein-kinin system: current and future pharmacological targets. *J Pharmacol Sci* **99**, 6-38 (2005).
39. Lieb, W. *et al.* Genome-wide meta-analyses of plasma renin activity and concentration reveal association with the kininogen 1 and prekallikrein genes. *Circ Cardiovasc Genet* **8**, 131-40 (2015).
40. (!!! INVALID CITATION !!!).
41. Battle, A., Brown, C.D., Engelhardt, B.E. & Montgomery, S.B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
42. Gamazon, E.R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* **50**, 956-967 (2018).
43. Fresquet, M. *et al.* PLA2R binds to the annexin A2-S100A10 complex in human podocytes. *Sci Rep* **7**, 6876 (2017).
44. Jiao, S., Zheng, X., Yang, X., Zhang, J. & Wang, L. Losartan inhibits STAT1 activation and protects human glomerular mesangial cells from angiotensin II induced premature senescence. *Can J Physiol Pharmacol* **90**, 89-98 (2012).
45. Lopez-Sanz, L. *et al.* SOCS1-targeted therapy ameliorates renal and vascular oxidative stress in diabetes via STAT1 and PI3K inhibition. *Lab Invest* (2018).
46. Eckardt, K.U. *et al.* Autosomal dominant tubulointerstitial kidney disease: diagnosis, classification, and management--A KDIGO consensus report. *Kidney Int* **88**, 676-83 (2015).

47. Gillies, C.E. *et al.* An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet* **103**, 232-244 (2018).
48. Dudley, A.J., Bleasby, K. & Brown, C.D. The organic cation transporter OCT2 mediates the uptake of beta-adrenoceptor antagonists across the apical membrane of renal LLC-PK(1) cell monolayers. *Br J Pharmacol* **131**, 71-9 (2000).
49. Filipski, K.K., Mathijssen, R.H., Mikkelsen, T.S., Schinkel, A.H. & Sparreboom, A. Contribution of organic cation transporter 2 (OCT2) to cisplatin-induced nephrotoxicity. *Clin Pharmacol Ther* **86**, 396-402 (2009).
50. Motohashi, H. & Inui, K. Organic cation transporter OCTs (SLC22) and MATEs (SLC47) in the human kidney. *AAPS J* **15**, 581-8 (2013).
51. Popejoy, A.B. & Fullerton, S.M. Genomics is failing on diversity. *Nature* **538**, 161-164 (2016).
52. Fuchsberger, C., Taliun, D., Pramstaller, P.P. & Pattaro, C. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics* **28**, 444-5 (2012).
53. Coresh, J. *et al.* Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA* **311**, 2518-2531 (2014).
54. Levey, A.S. *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* **150**, 604-12 (2009).
55. Pattaro, C. *et al.* Estimating the glomerular filtration rate in the general population using different equations: effects on classification and association. *Nephron Clin Pract* **123**, 102-11 (2013).
56. Schwartz, G.J. *et al.* Improved equations estimating GFR in children with chronic kidney disease using an immunonephelometric determination of cystatin C. *Kidney Int* **82**, 445-53 (2012).
57. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
58. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
59. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).
60. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMC glmm R Package. *J Stat Softw* **33**, 1-22 (2010).
61. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J Transl Med* **13**, 348 (2015).
62. Noce, D. *et al.* Sequential recruitment of study participants may inflate genetic heritability estimates. *Hum Genet* **136**, 743-757 (2017).
63. Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
64. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
65. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**, e93766 (2014).
66. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
67. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat Genet* **50**, 1593-1599 (2018).
68. Kottgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* **45**, 145-54 (2013).

69. Fehrmann, R.S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-25 (2015).
70. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
71. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-6 (2007).
72. Hoppmann, A.S., Schlosser, P., Backofen, R., Lausch, E. & Kottgen, A. GenToS: Use of Orthologous Gene Information to Prioritize Signals from Human GWAS. *PLoS One* **11**, e0162466 (2016).
73. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
74. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
75. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* **33**, 79-86 (2009).
76. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334-6 (2015).
77. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
78. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-70 (2010).
79. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
80. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
81. Fehrmann, R.S. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet* **7**, e1002197 (2011).
82. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243 (2013).
83. Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* **18**, 16 (2017).
84. Kirsten, H. *et al.* Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding locidagger. *Hum Mol Genet* **24**, 4746-63 (2015).
85. Beutner, F. *et al.* Rationale and design of the Leipzig (LIFE) Heart Study: phenotyping and cardiovascular characteristics of patients with coronary artery disease. *PLoS One* **6**, e29070 (2011).
86. Loeffler, M. *et al.* The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health* **15**, 691 (2015).
87. Eckardt, K.U. *et al.* The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol Dial Transplant* **27**, 1454-60 (2012).